PB89-233506

Air Quality Simulation Model Performance for One-Hour Averages

Systems Applications, Inc., San Rafael, CA

Prepared for:

Atmospheric Research & Exposure Assessment Laboratory
Research Triangle Park, NC

Aug 89

U.S. Department of Commerce
National Technical Information Service
NTIS

PB89-233506

EPA/600/3-89/071
August 1989

AIR QUALITY SIMULATION MODEL PERFORMANCE FOR ONE-HOUR AVERAGES

by

G. E. Moore
T. E. Stoeckenius
Systems Applications, Inc.
101 Lucas Valley Road
San Rafael, California 94903

and

Steve Hanna
Dave Strimitas
Sigma Research Corp.
394 Lowell St.
Lexington, MA

Contract No. 68-02-4352 Option I
Work Assignment 2-13

Project Officer

D. Bruce Turner
Meteorology and Assessment Division
Atmospheric Research & Exposure Assessment Laboratory
Research Triangle Park, NC 27711

Atmospheric Research & Exposure Assessment Laboratory
OFFICE OF RESEARCH AND DEVELOPMENT
U. S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/3-89/071	2.	3. RECIPIENT'S ACCESSION NO. PB89-233506/AS
4. TITLE AND SUBTITLE AIR QUALITY SIMULATION MODEL PERFORMANCE FOR ONE-HOUR AVERAGES		5. REPORT DATE August 1989
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) [1]G.E. Moore, T.E. Stoeckenius [2]Steve Hanna and Dave Strimitas		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS [1]Systems Applications, Inc - 101 Lucas Valley Rd San Rafael, California 94903 [2]Sigma Research Corp. - 394 Lowell St. Lexington, MA		10. PROGRAM ELEMENT NO. A101/B/69/08 7031-01 (FY-88) 11. CONTRACT/GRANT NO. 68-02-4352
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Research & Exposure Assessment Laboratory Office of Research and Development U. S. Environmental Protection Agency Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA/600/09

15. SUPPLEMENTARY NOTES

16. ABSTRACT

If a one-hour standard for sulfur dioxide were promulgated, air quality dispersion modeling in the vicinity of major point sources would be an important air quality management tool. Would currently available dispersion models be suitable for use in demonstrating attainment of such a standard in the vicinity of large, elevated, buoyant point sources such as utility power plants? The results summarized in this report suggest that using these models in connection with an hourly average standard does not present the regulatory community with any significant additional uncertainties that are not already being dealt with in connection with the current 3- and 24-hour standards. Predictions of peak hourly average concentrations unmatched in time or location are slightly more conservative on average than corresponding predictions of 3- and 24-hour averages. Although certain conditions not specifically treated by the models, such as inversion break-up fumigation, are known to occur and to have the potential for producing high ground-level impacts, these conditions have been only rarely observed in model evaluation studies and do not seem to contribute to the bulk of the 50 highest concentrations. In fact, nearly all of the highest concentrations observed in the evaluation studies examined occurred under convective conditions, which are treated by the existing models. Peak model predictions under such conditions match reasonably well with the highest observed concentrations, although large prediction errors can occur in any given hour. Consideration of situations involving impacts on elevated terrain or building downwash phenomena have been explicitly excluded. This review of the results of model evaluation studies was restricted to some fairly simple analyses. Additional analyses are recommended. Recommendations to improve model formulations are made regarding the treatment of plume behavior in the vicinity of the mixing height and the simulation of plume dispersion within the mixed layer during convective conditions.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 150
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE A07

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

The information in this document has been funded by the United States Environmental Protection Agency under Contract No. 68-02-4352 Option I, Work Assignment 2-13 to Systems Applications, Inc. and Sigma Research Corp. It has been subject to the Agency's peer and administrative review, and it has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

If a one-hour standard for sulfur dioxide were promulgated, air quality dispersion modeling in the vicinity of major point sources would be an important air quality management tool. Would currently available dispersion models be suitable for use in demonstrating attainment of such a standard in the vicinity of large, elevated, buoyant point sources such as utility power plants? The results summarized in this report suggest that using these models in connection with an hourly average standard does not present the regulatory community with any significant additional uncertainties that are not already being dealt with in connection with the current 3- and 24-hour standards. Predictions of peak hourly average concentrations unmatched in time or location are slightly more conservative on average than corresponding predictions of 3- and 24-hour averages. Although certain conditions not specifically treated by the models, such as inversion break-up fumigation, are known to occur and to have the potential for producing high ground-level impacts, these conditions have been only rarely observed in model evaluation studies and do not seem to contribute to the bulk of the 50 highest concentrations. In fact, nearly all of the highest concentrations observed in the evaluation studies examined occurred under convective conditions, which are treated by the existing models. Peak model predictions under such conditions match reasonably well with the highest observed concentrations, although large prediction errors can occur in any given hour. Consideration of situations involving impacts on elevated terrain or building downwash phenomena have been explicitly excluded. This review of the results of model evaluation studies was restricted to some fairly simple analyses. Additional analyses are recommended. Recommendations to improve model formulations are made regarding the treatment of plume behavior in the vicinity of the mixing height and the simulation of plume dispersion within the mixed layer during convective conditions.

CONTENTS

Notice ii
Abstract iii
Figures vii
Tables ix
Acknowledgments xii

1 INTRODUCTION 1

Background to the Study 1
Scope of the Study 3
Report Organization 3

2 REVIEW OF PAST MODEL PERFORMANCE EVALUATION STUDIES 5

Statistical Measure of Interest 6
Major Model Performance Studies 6
  EPA Power Plant Study 10
  Maryland Power Plant Siting Program 11
  API Tracer Evaluation Study 13
  EPRI Plume Model Validation and Development Study 17
  EPA Rural Model Evaluation Study 20
  EPA Urban Model Evaluation Study 25
  Other Model Performance Studies 28
Selection of Studies for Further Model Performance 30

3 REVIEW OF MODEL FORMULATIONS 35

Introduction 35
Vertical Structure 37
  CRSTER 37
  TEM 38
  MPTER 38
  RTDM 38
  SHORTZ 39
  PPSP 39
  HPDM 40
Plume Rise and Dispersion 41
  CRSTER 41
  TEM 42
  MPTER 42
  RTDM 42

Preceding page blank

SHORTZ 43
PPSP 44
HPDM 45

Terrain Formulations 46

CRSTER 46
MPTER 47
SHORTZ 47
RTDM 47

Treatment of Calm Conditions 48

4 OVERALL PERFORMANCE OF THE MODELS 49

The Data Sets 49

The Highest Observations 52

Largest Overpredictions 55

Largest Underpredictions 55

Meteorology Associated with Highest Concentrations 62

5 MODEL PERFORMANCE AS A FUNCTION OF INPUT DATA 69

Analysis of Hourly SO_2 Concentration Data--The Kincaid Data 69

Scatterplots and Correlational Analysis 74

Cluster Analysis of the Largest Residuals 81

Discriminant Analysis of the Residuals 90

6 SYNTHESIS, SUMMARY, AND CONCLUSIONS 95

Model Performance: Ability to Predict Peak Concentrations 96

Model Performance Under Specific Dispersion Conditions 105

Summary and Conclusions 118

Recommendations 122

References 125

Appendix: Maximum One-Hour SO_2 Concentrations Analyzed by Scaling Procedures

FIGURES

Number Page

1-1 Various types of plume behavior leading to maximum ground level concentrations 2

2-1 Comparison between measured and predicted ground-level SO2 concentrations for the (a) PPSP and (b) CRSTER models 14

2-2 Maximum 1-hour SO2 concentration for each hour categorized by stability group 19

2-3 Five highest 1-hour average observed and predicted χ/Q values (10-9s/m3) unpaired in time or location--all data in the common data set (93) hours) 21

4-1 A scatter plot of predictions and observations for the 50 hours when the highest concentrations were observed at Kincaid 53

4-2 A scatter plot of predictions and observations for the 50 hours when the highest concentrations were observed at Clifty Creek. 54

4-3 A scatter plot of the 50 largest overpredictions and underpredictions of normalized SF6 made by CEQM at Kincaid 58

4-4 A scatter plot of observations and predictions for the 50 hours with the highest observed concentrations. 59

4-5 A scatter plot of the largest 7 overpredictions by TEM of normalized SF6 concentrations at Kincaid. 60

4-6 Values of buoyancy flux and mixing height associated with underpredictions made by CEQM at the Bull Run site 63

5-1 Scatter plots of residuals versus various model inputs and stability and scaling parameters for CEQM at Bull Run 77

5-2 Scatter plots of residuals versus various model inputs and stability and scaling parameters for CEQM at Kincaid (SF6) 78

5-3 Scatter plots of residuals versus various model inputs and stability and scaling parameters for CRSTER at Clifty Creek 79

FIGURES (Continued)

Number Page

5-4 Scatter plots of residuals versus various model inputs and stability and scaling parameters for PPSP at the Maryland PPSP sites.. 80

5-5 Scatter plots of residuals (observation minus prediction) of the 50 highest concentrations versus model inputs and stability and scaling parameters for RTDM at Bull Run........ 82

5-6 A scatter plot of case cluster number versus the first two canonical variables. The plot is for CEQM at Bull Run....... 92

5-7 A scatter plot of case cluster number versus the first two canonical variables. The plot is for PPSP at the Maryland PPSP sites.. 94

6-1 Dependence of model bias on averaging time in the EPA Rural Study (Clifty Creek 1975 and 1976 data).................... 101

6-2 Dependence of time correlation on averaging time for EPRI-PMV Kincaid study (maximum SO2 concentrations; predictions and observations paired in time only).......................... 102

6-3 Dependence of model bias on atmospheric stability class in the EPA Rural Study (Clifty Creek 1975 data)............... 114

6-4 Dependence of model bias on atmospheric stability class in the EPA Urban Study (St. Louis 1976 data). 115

6-5 Dependence of model bias on atmospheric stability class in the API Tracer Study... 116

TABLES

Number Page

2-1 Major model evaluation studies reviewed 7

2-2 Miscellaneous model evaluation studies reviewed 8

2-3 Models evaluated by the major evaluation studies 9

2-4 Meteorologically stratified residuals (μg/m3) for observations minus predictions for the Clifty Creek Ridge monitor in 1975 for hours when maximum impact is likely 12

2-5 The ratios of the average 10 highest predictions and average 10 highest observations from the API model performance evaluations using observations from elevated tracer releases 16

2-6 Comparison of 25 highest 1-hour average observed and predicted SO2 concentration values (ppb) at Kincaid, Illinois, that are unpaired in time and location 18

2-7 Highest 1-hour average observed and predicted χg/Q (10-9 s/m3) paired in time--all data in common data set 22

2-8 Average residuals (observations minus predictions) of the 25 highest observed 1-hour SO2 concentrations (μg/m3) for the performance evaluation of models using the 1975-1976 Clifty Creek data sets 24

2-9 Average residuals (observations minus predictions) of the 25 highest observed and predicted unpaired 1-hour SO2 concentrations (μg/m3) for selected data bases 26

2-10 Performance statistics from the EPA urban model evaluation study for the highest 25 unpaired 1-hour SO2 concentration observations and predictions 27

2-11 A summary of RAM performance evaluation studies 29

2-12 Data sets selected for further study 31

2-13 Data sets rejected for further study 33

2-14 A summary of the meteorological data reported during the model evaluation exercises selected for further study 34

TABLES (Continued)

Number Page

3-1 Air quality models and data sets included in model evaluation studies 36

3-2 Components particularly important to certain meteorological categories 37

4-1 Summary of statistics for time-matched observations and predictions for the 50 hours with the highest observed concentrations 56

4-2 Summary of statistics for time-matched observations and predictions for hours with the largest overpredictions 57

4-3 Summary of statistics for time-matched observations and predictions for hours with the largest underpredictions 61

4-4 Ranges for meteorological variables that occur most frequently during the 50 hours with the highest concentrations 64

4-5 Ranges for meteorological variables that occur most frequently during hours with the largest underpredictions 65

4-6 Ranges for meteorological variables that occur most frequently during hours with the largest overpredictions 67

5-1 Data for top 25 hourly SO2 concentrations at Kincaid for developmental data set 71

5-2 Data for top 25 hourly SO2 concentrations at Kincaid for evaluation data set 72

5-3 Variation of maximum hourly SO2 concentrations with U and L at Kincaid 73

5-4 Correlation coefficients of predicted minus observed concentration residuals versus model input parameters for the 50 hours with the highest observed concentrations 75

5-5 Cluster standardized means from the 50 hours with the largest magnitude residuals for CEQM at Bull Run using routine model input data 84

TABLES (Continued)

Number Page

5-6 Cluster standardized means from the 50 hours with the largest magnitude residuals for CEQM at Bull Run using convective scaling variables........................ 84

5-7 Cluster standardized means from the 50 hours with the largest magnitude residuals for CEQM at Kincaid (SF6).. 86

5-8 Cluster standardized means from the 50 hours with the largest magnitude residuals for CRSTER at Clifty Creek............. 86

5-9 Cluster standardized means from the 50 hours with the largest magnitude residuals for the PPSP model during the Maryland PPSP.. 88

5-10 Cluster standardized means from the 50 hours with the largest magnitude residuals for RTDM at Bull Run........... 89

5-11 A summary of the canonical coefficients (weights) for the two canonical variables that explain most of the variance within the clusters................................. 91

6-1 Summary of model performance for peak concentrations unpaired in time or location.............................. 97

6-2 Summary of model performance for the hours with the 50 highest observed concentrations: Observed and predicted concentrations matched in time but not location........... 104

6-3 Description of several generalized plume behaviors potentially associated with high ground-level concentrations and/or large modeling errors.................................... 107

ACKNOWLEDGEMENTS

We wish to thank Mr. John Irwin, Mr. Thomas E. Pierce, and our Project Officer, Mr. D. Bruce Turner, of the U.S. Environmental Protection Agency for their helpful comments on the draft version of this report. We also wish to thank Dr. C.S. Burton, Mr. Stan Hayes, and Mr. Len Milich of Systems Applications, Inc. for their assistance and words of encouragement. Finally, we express our sincere appreciation to our editor, Mr. Howard Beckman, for his untiring efforts to mold our awkward prose into readable form.

1 INTRODUCTION

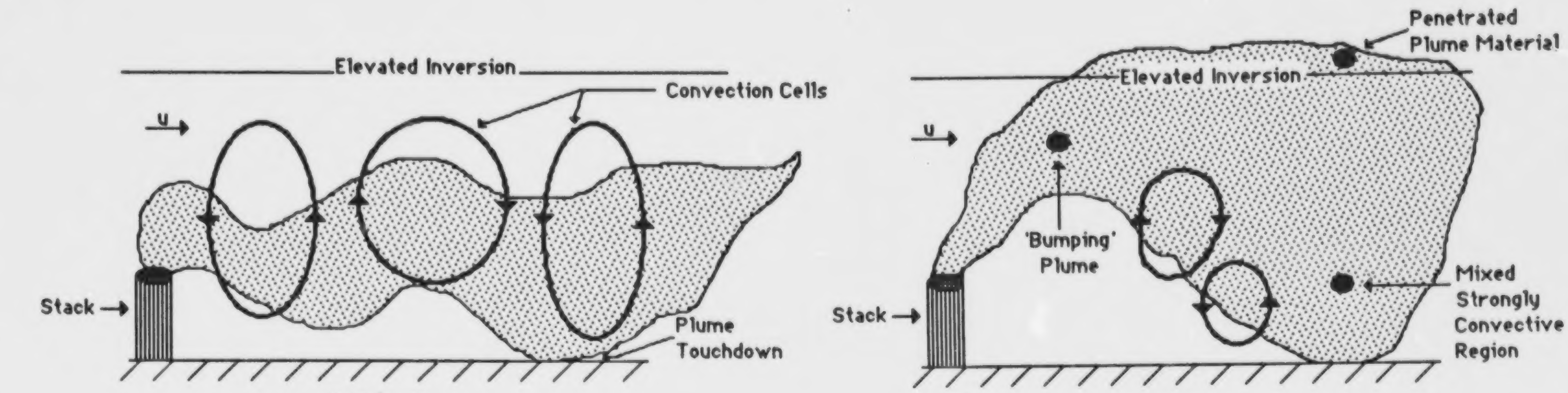
In connection with its review of the National Ambient Air Quality Standard (NAAQS) for sulfur dioxide, the EPA has been considering the possibility of setting a one-hour standard to protect certain sensitive members of the population against exposure to short-term, peak SO_2 concentrations. If such a standard were to be promulgated, air quality dispersion modeling of hourly average SO_2 concentrations in the vicinity of major point sources would become an important air quality management tool. Although a number of dispersion models that calculate hourly average concentrations in the vicinity of such sources have been approved by the EPA, current regulatory applications of these models only require accurate estimation of the highest and "high second high" 3-hour and 24-hour average concentrations. Promulgation of a one-hour standard (which may be stated in an expected exceedance form) would impose the requirement that these models accurately estimate the highest and high second high as well as the third, fourth, etc., highest hourly average concentrations.

These new requirements bring up the question of just how suitable currently available dispersion models are for this purpose. In other words, how well do the models predict peak hourly average concentrations? Are there particular meteorological conditions under which the models do especially well or poorly? Are certain meteorological regimes (i.e., fumigation), which are not explicitly treated by the models, responsible for high observed hourly concentrations? This report represents an attempt to answer some of these questions.

BACKGROUND TO THE STUDY

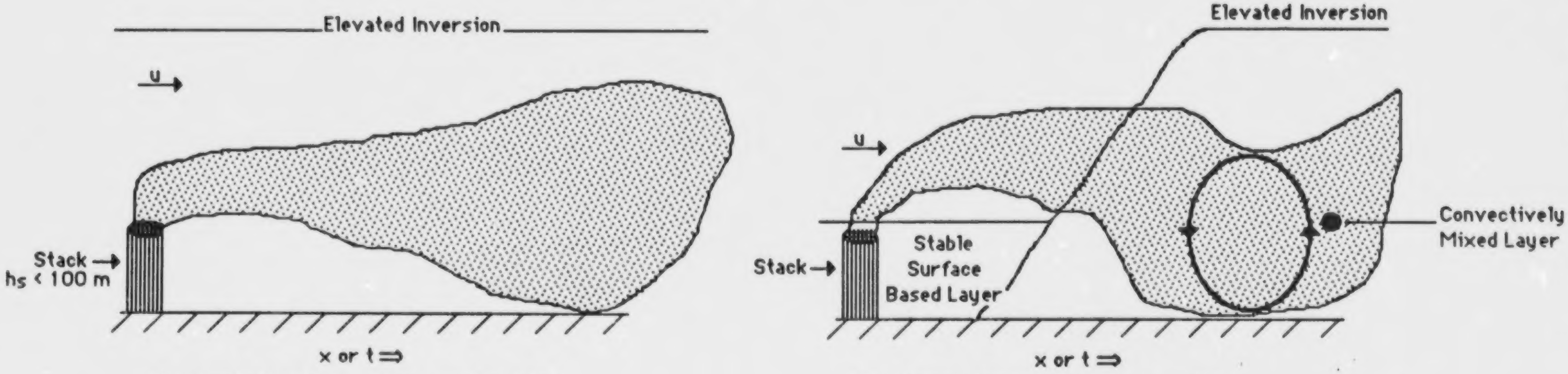
The performance of the U.S. EPA guideline models (and candidates) has been evaluated in a number of studies. Short-term guideline models for large SO_2 point sources use hour-by-hour emissions and meteorological data to produce hour-by-hour predictions of ground-level concentration.

It has long been recognized by air quality modelers that certain types of buoyant point source plume behaviors are responsible for the highest concentrations and that the behaviors are associated with specific dispersion conditions. Among the most commonly identified conditions are convective looping, various plume trapping phenomena, and high-wind fumigation due to low plume rise. Figure 1-1 depicts some of these plume behaviors.


Elevated Inversion
Convection Cells
u
Stack
Plume Touchdown
x or t ⇒
Penetrated Plume Material
Elevated Inversion
u
'Bumping' Plume
Stack
Mixed Strongly Convective Region
x or t ⇒

(a) Plume looping under convective conditions

(c) Plume "bumping" and convective looping and fumigation


Elevated Inversion
u
Stack
$h_s < 100$ m
x or t ⇒
Elevated Inversion
u
Stack
Stable Surface Based Layer
Convectively Mixed Layer
x or t ⇒

(b) Limited plume rise under high wind, neutral conditions

(d) Plume fumigation as a stable coning/fanning plume encounters a thermal internal boundary layer or a growing surface based mixed layer

FIGURE 1-1. Various types of plume behavior leading to maximum ground level concentrations.

SCOPE OF THE STUDY

In the study reported here we reviewed the published evaluations of several short-term air quality models. We obtained some of the data sets from these evaluations to examine the performance of various models for the hours with the largest 50 observations and the 50 largest magnitude residuals. Extensive meteorological data were used to stratify these two concentration data sets and model performance statistics. An attempt was made to determine why the models performed as they did by analyzing the available aerometric and meteorological data. To do this we made use of recent developments in convective boundary layer (CBL) dispersion theory to provide insight into the occurrence and characteristics of the highest 1-hour SO_2 concentrations.

Because a 1-hour SO_2 NAAQS is being considered, the primary focus of this study was on SO_2 and tracer experiments within 50 km of large elevated point sources with buoyant emissions that are typical of fossil-fueled power plants. In order to avoid site-specific complications that may confuse the performance evaluation results, we concentrated our analysis on model applications to flat terrain where there was also no interference from nearby sources.

An effort was also made to define dispersion conditions that produce the highest 1-hour concentrations and to stratify the paired observed-predicted data by these conditions. This was done in order to explain the reasons for inaccurate model results under each condition.

REPORT ORGANIZATION

The report describes the findings of the six tasks originally outlined in the scope of work for this project.

(1) Review of model performance evaluation and observed-predicted concentration data sets, and selection of observed-predicted concentration data sets and models for further study of model performance for one-hour averaging periods.

(2) Identification of dispersion conditions when the highest observed concentrations occurred, and when the largest observed minus predicted concentrations (residuals) occurred.

(3) A review of the overall performance of several models for the hours with the highest observed concentrations or the largest over- or underpredicted concentrations.

(4) Description of the formulation of selected models and an assessment of the adequacy of the formulation for dealing with the meteorological conditions identified in Task 2.

(5) Analysis of model performance under each of the meteorological conditions identified in Task 2.

(6) Summary and synthesis of the findings of Task 5.

The work of these tasks is reported in five chapters. In Chapter 2 we review the selection of appropriate model evaluation studies and specific data sets for further study. In Chapter 3 we review the formulation of the models and assess how the models might be expected to perform during dispersion conditions when the highest concentrations occur. In Chapter 4 we summarize the results of the performance evaluations for each of the models and describe the data bases used in the performance evaluations in this study. In Chapter 5 we analyze model performance using a number of stratifications by various meteorological conditions. Model performance statistics, such as bias, are examined to determine why a model performed as it did. In Chapter 6 we synthesize the results from the various model performance studies and recommend improvements to the models that will lead to more accurate one-hour concentrations.

Readers who wish to skip the technical details of our analysis may refer directly to Chapter 6. The final section of this chapter provides a concise summary of our results.

2 REVIEW OF PAST MODEL PERFORMANCE EVALUATION STUDIES

This section has two purposes. The first is to review published evaluations of the ability of selected models to calculate hourly average concentrations. The second is to select from these studies appropriate data sets for further study of hours when worst-case dispersion conditions lead to maximum observed or predicted ground-level concentrations. Since the NAAQS for SO_2 under consideration is a 1-hour standard, the focus of this review is on the performance of models at sites within 50 km of large point sources of SO_2. The emissions inventories for SO_2 point sources are usually dominated by large power plants and industrial complexes. These types of emissions are generally elevated and buoyant. Therefore we reviewed only models that will treat short-term emissions from such sources (we did not review models that produce only annual averages).

Short-term Gaussian point source models are numerous, and most follow a common set of dispersion and plume rise formulas similar to those found in the U.S. EPA guideline models (EPA, 1986). However, most of these models were developed for specific applications and are neither EPA guideline models nor recognized alternative models. We did not consider such models since they are not of primary interest to the US EPA.

Of particular interest were evaluations where relationships between dispersion conditions and the highest concentrations could be identified and statistically quantified as correlations, regressions, clusters, and so on. Therefore we selected evaluation studies that would not interfere with such analyses. We ignored evaluations of model applications in situations where the plume can impinge on monitors in elevated terrain and in areas where significant point sources from other power plants might interfere with a performance evaluation by providing at times an elevated background concentration at some receptors. Other studies that were not considered are those where meteorological phenomena lead to inhomogeneous and time-varying dispersion conditions, e.g., land-sea breeze regimes and slope flows.

We have organized our review by major evaluation programs wherever possible, and by type of environment (rural, urban, and complex terrain). Of the large number of evaluation studies, some have produced interesting results but most use relatively limited data and often nonstandard statistical measures to make a specific point. As a result we have not spent much time reviewing such studies, either trying to reformulate statistics or to characterize findings, unless they relate specifically to performance under dispersion conditions that lead to extreme observed ground-level concentrations.

STATISTICAL MEASURES OF INTEREST

The first and second highest and the N highest observed 1-hour concentrations are of major interest to this study. We are also interested in hours when dispersion meteorology produces the N highest observations. Pairing of observations and predictions by location is not particularly relevant to our purpose, as it would be for PSD considerations, and as was done in the review of model performance by Hayes and Moore (1986). Thus we will pay particular attention to

> Highest observed and predicted concentrations paired only by rank, and not by location or time
>
> Highest observed and predicted concentrations paired by hour (one pair per hour)

Prior to the 1980s a number of statistical measures were created to quantify particular factors of model performance. Many of these measures were either statistically ill conditioned, or else had multiple interpretations. During the early 1980s several workshops, such as that reported by Fox (1981), led to a more or less standard set of statistical measures. A few additions have been introduced since then. The most used additional statistic is the fractional difference measure, $2(O - P)/(O + P)$, introduced by Hayes (1979) and by Cox (1983), which avoids some of the unpleasant characteristics of ratio measures, such as that of the ratio of predictions to observations (P/O). We will focus on residual measures of bias and precision, correlations, and some ratio measures such as the fractional difference in our review. We are aware that the sample sizes of the observation-prediction data sets vary greatly, and the uncertainties of the statistics certainly vary as well. We will report confidence intervals wherever they exist in the evaluation studies we review, but we have not gone out of our way to make independent estimates of uncertainty, however badly they might be needed.

MAJOR MODEL PERFORMANCE STUDIES

We selected six major studies from among the rural and urban model evaluations of point source models examined. These are listed and briefly described in Table 2-1. A number of limited studies either involved complex terrain, or else were RAM validations in urban areas and as such were not of direct interest because of the presence of background concentrations. Table 2-2 presents a summary of these miscellaneous evaluation studies that were closest to being useful for the present study.

The models that were evaluated in the major studies are listed in Table 2-3. As this table shows, most of the studies evaluated CRSTER/MPTER or some equivalent. All of the EPA guideline models share common features, among which are:

TABLE 2-1. Major model evaluation studies reviewed.

Study/References	Description
EPA POWER PLANT STUDY Mills and Stern, 1975 Mills and Record, 1975 Mills et al., 1979 Turner and Irwin, 1982	Collection of three EPA evaluation studies at the Muskingum, Clifty Creek, Tanner's Creek, Gavin-Kyger-Sporn, Stuart, Philo, and Canal power plants
MARYLAND POWER PLANT SITING PROGRAM Weil and Brower, 1982 Weil and Brower, 1984 Weil, 1979	Intensive collection of observations of plume parameters at the Dickerson, Chalk Point, and Morgan Town power plants were used to develop and evaluate models
API TRACER MODEL EVALUATION STUDY Londergan et al., 1980 Hellums, 1982	An evaluation study of rural and urban models used a dozen tracer studies with surface and elevated releases
EPRI PLUME MODEL VALIDATION AND DEVELOPMENT (PMV&D) PROGRAM Reynolds et al., 1984 Reynolds et al., 1985	An intensive measurement and evaluation program was carried out at a flat site (Kincaid) and a site in rolling terrain (Bull Run)
EPA RURAL MODEL EVALUATION STUDY Londergan et al., 1982 Smith, 1984 Irwin and Smith, 1984 Cox and Moss, 1985 Cox et al., 1987 Cox et al., 1986	The Clifty Creek, Muskingum, Kincaid and Paradise data sets were used to evaluate up to 10 rural point source models
EPA URBAN MODEL EVALUATION STUDY Londergan et al., 1983 Ruff, 1980 Rao et al., 1985 Turner et al., 1985	Two models that predict hourly concentrations were exercised and evaluated using the St. Louis RAPS data base consisting of hourly SO_2 concentrations at 13 sites

TABLE 2-2. Miscellaneous model evaluation studies reviewed (this does not include RAM studies at various cities).

Study/References	Description
MICHIGAN DOWNWASH STUDY Touma et al., 1981	This study examined the effects of adding partial plume penetration and building downwash to RAMR. A large power plant in Michigan was selected for the validation data base
WIDOWS CREEK EVALUATION STUDY Lott, 1984	This was a case study of maximum concentrations occurring on elevated terrain around the TVA power plant at Widow's Creek, during July 1978 to February 1979
VISTTA FIELD STUDY EVALUATION White et al., 1985	An evaluation of cross plume integrals of concentration of NO_2 and TSP for 4 visibility models for observations around the Navajo Power Plant in Arizona (models were PLUVUE, ERT, PHOENIX, and LOS ALAMOS)
ALUMINUM REDUCTION PLANT EVALUATION STUDY Scire and Schulman, 1981 McTaggart, 1981	SF_6 tracer was used to evaluate BLP and ISC(ST) performance at an aluminum reduction plant in Arkansas and one in Pittsburg, Pennsylvania

TABLE 2-3. Models evaluated by the major evaluation studies (listed in Table 2-1).

Study	Models (hourly predictions)
EPA Power Plant Study	CRSTER
Maryland Power Plant Siting Program	CRSTER/MPTER and PPSP
API Tracer Model Evaluation	PTMTP, PAL, CRSTER, RAMR, RAM
EPRI Plume Model Validation and Development (PMV&D) Program	
Kincaid	CEQM(CRSTER), TEM, MSDM, HPDM*
Bull Run	CEQM(CRSTER), SHORTZ, RTDM, HPDM*
EPA Rural Model Evaluation Study	MPTER/CRSTER, PLUME5, MPSDM, COMPTER, SCRSTER, 3141/4141, TEM-8A, MULTIMAX Note: CRSTER/MPTER, TEM-8A, MPSDM, PPSP were evaluated for other data bases
EPA Urban Model Evaluation Study	RAM, TEM-8A

* Preliminary results only.

Pasquill-Gifford dispersion coefficients
Turner stability classification
Briggs plume rise formula
Power law wind speed extrapolation
Ground/inversion reflection
RAMMET (CRSTER) meteorological preprocessor input

Only two of the models--RTDM and PPSP--use turbulence measures to determine dispersion rates, as recommended by Hanna and others (1977). Of these two models, only PPSP has special provisions for modeling concentrations under convectively unstable conditions. A third model, the Hybrid Plume Dispersion Model (HPDM), developed by Paine and Hanna (1986), also uses turbulence measures and has a convective instability dispersion algorithm. These differences in model formulation will be discussed in Section 3.

It should be noted that the models used in the studies we selected have undergone some revision since the studies were conducted. The evaluated versions are not necessarily the same as the present UNAMAP 6 versions. Options have been added to models and default values for certain model parameters have been revised. Wherever possible we will try to identify a version of the code. The PPSP and HPDM are relatively new models and the data sets used for development are part of or else related to the model performance evaluation data set. Other models, such as CRSTER, TEM, and RTDM, have gone through several revisions that need to be identified.

EPA Power Plant Study

The EPA Power Plant study comprises three model evaluation studies. The air quality and meteorological data from six power plant sites in the Midwest were used in the evaluation studies--Muskingum, Stuart, Philo, Clifty Creek, Tanner's Creek, and Gavin-Kyger-Sporn. A seventh site, Canal, was located on a peninsula on the East Coast. A review of the reports revealed the following.

> The evaluations for the Muskingum, Stuart, Philo, and Canal power plants were severely limited in extent, as evidenced by the choice of statistics for the evaluations.
>
> One-hour evaluation results were not available for the Tanner's Creek or Gavin-Kyger-Sporn sites (only 3- and 24-hour averages were available)
>
> Only the Clifty Creek evaluation had performance statistics broken down by meteorology (mixing height, PGT stability, and wind direction)

The basic finding obtained from the Stuart, Muskingum, Philo and Canal studies is that CRSTER (the 1973 UNAMAP 1 version) overpredicts the highest and second highest concentrations. Another finding was the crossover in the cumulative distribution function so that at lower concentrations CRSTER underpredicted concentrations. In this study the 99th percentile predictions are generally lower than the 99th percentile observations.

When converted to the fractional difference, 2(O - P)/(P + O), the results range from near zero for the Canal site to -1.0 for the Philo site. The large overpredictions at the Philo site occur because some of the differences in elevation between monitors and the stack base are greater than 0.7. Turner and Irwin (1982) point out that when this condition occurs one can generally expect to see overpredictions because of partial plume impingement on the terrain.

The relation of peak concentration to averaging period was briefly addressed by comparing the maximum 1-hour and 3-hour average concentrations with the 24-hour maximums. Generally the 1-hour concentrations were twice those of the 3-hour average and 7-16 times those of the 24-hour averages.

Studies of the Clifty Creek site conducted by Mills and co-workers (1979) provide a much richer data set for study. Statistical measures were computed for each of the six monitors. In addition, a modification to CRSTER, based on some of the recommendations made by Irwin (1979), was evaluated to see if such modifications actually improved model performance. For one of the monitors the bias is a function of mixing height range and PGT stability class (Table 2-4). Both versions of CRSTER possess a slight overall tendency to underpredict at this monitor. The greatest overprediction occurs when mixing heights are high and conditions are unstable. A second maximum in the frequency of overpredictions occurs when mixing heights are moderately high and conditions are slightly stable. The greatest underprediction occurs when the mixing height is low and conditions are stable; a second combination that seems to lead to underpredicition is neutral stability conditions with high mixing height.

The CRSTER modifications resulted in larger overpredictions or changed some station statistics from underpredictions to overpredictions. At the same time, however, the correlation coefficients ranged from 0.31 to 0.61 for the unmodified CRSTER and were increased to 0.39 to 0.76 for the modified version. So while the improvements resulted in better hour-by-hour association, the bias statistics were worsened.

Maryland Power Plant Siting Program

The Maryland Power Plant Siting Program (PPSP) made intensive measurements during unstable and neutral conditions at three power plants in Maryland (Dickerson,

TABLE 2-4. Meteorologically stratified residuals (μg/m^3) for observations minus predictions for the Clifty Creek Ridge monitor in 1975 for hours when maximum impact is likely. Numbers in parentheses are numbers of observations. The number of observations is the same for both CRSTER versions. (Source: Mills et al., 1979)

Stability	max ←		Mixing Height Categories		→	min	
Class	6	5	4	3	2	1	All
			CRSTER (Turner)				
A	-56.7 (6)	-116 (9)	-- (0)	-- (0)	* (1)	-- (0)	-84.1 (16)
B	-39.0 (31)	-38.0 (54)	23.6 (19)	* (2)	* (2)	* (1)	-24.1 (109)
C	-14.8 (69)	0.4 (120)	3.4 (74)	* (18)	* (6)	* (2)	-1.66 (289)
D	16.0 (137)	17.4 (573)	8.4 (498)	-2.8 (148)	* (39)	* (32)	12.9 (1427)
E	-10.9 (71)	-4.5 (170)	-49.0 (77)	-5.5 (12)	-23.8 (12)	-9.4 (13)	-16.3 (355)
F	7.8 (128)	2.3 (113)	1.7 (46)	* (8)	-13.3 (4)	10.3 (8)	+4.5 (307)
G	* (54)	* (65)	* (21)	* (1)	-- (0)	-- (0)	* (141)
All	0.8 (496)	6.0 (1104)	3.2 (735)	0.2 (189)	6.8 (64)	28.0 (56)	4.3 (2644)
			CRSTER (Irwin)				
A	-56.6	-108.	--	--	*	--	-79.6
B	-36.6	-20.2	42.8	*	*	*	-11.4
C	11.1	5.6	23.3	*	*	*	6.4
D	14.6	14.7	7.9	22.4	*	*	14.1
E	-11.4	-22.4	-76.3	-5.6	-24.3	7.4	-30.3
F	8.1	3.0	1.5	*	-9.9	13.8	4.9
G	*	*	*	*	--	--	*
All	1.2	3.4	2.5	19.5	6.9	32.4	4.6

* Neither CRSTER (Turner) nor CRSTER (Irwin) predicted a significant impact during hours with possible impact (wind direction within 30° of the source-receptor centerline).

Chalk Point, and Morgantown). Remote sensing was used to ascertain plume parameters aloft for 30-60 hours at each site. Two models were evaluated: the CRSTER model (UNAMAP 3 version) and a model developed from observations and known as the PPSP model. The PPSP model was developed by Weil (1980) as a continuation of his work on simple dispersion models for convectively unstable conditions.

The principal findings of the evaluation exercise were the following:

> The CRSTER model produced a substantial number of zero predictions when a moderate concentration was observed.
>
> The ratio P/O showed far greater scatter for CRSTER than for PPSP, particularly for underpredictions (P/O < 1).
>
> The correlation for CRSTER results was poor (-0.26) due to the zero and near-zero predictions. The PPSP model, which did not produce such zero predictions, had a correlation of 0.76.
>
> CRSTER generally underpredicted, with particularly serious underpredictions for neutral conditions. The PPSP model always overpredicted.
>
> CRSTER predictions produced a mean square error (MSE) that was four times greater than that produced by the PPSP predictions.

The interpretation of the results was somewhat confused by the use of a log transformation of the concentrations.

The primary differences in the performance of the two models are summarized in Figure 2-1. The poor CRSTER performance seems to be due in part to a number of zero ground-level concentration predictions. These occur because when CRSTER calculates the plume centerline to be above the mixing height, it assumes that the whole plume is lofting and none of the plume material reaches the ground. Without a partial plume penetration algorithm, CRSTER produces an "all-or-nothing" concentration prediction. Another reason seems to be the use of too small a vertical dispersion coefficient during neutral conditions, which results in nonzero but very small predictions.

API Tracer Evaluation Study

The American Petroleum Institute (API) commissioned a study to evaluate the performance of a number of regulatory air quality models using data from many tracer studies that had been used to study dispersion from surface and elevated tracer releases. The elevated releases included the following:

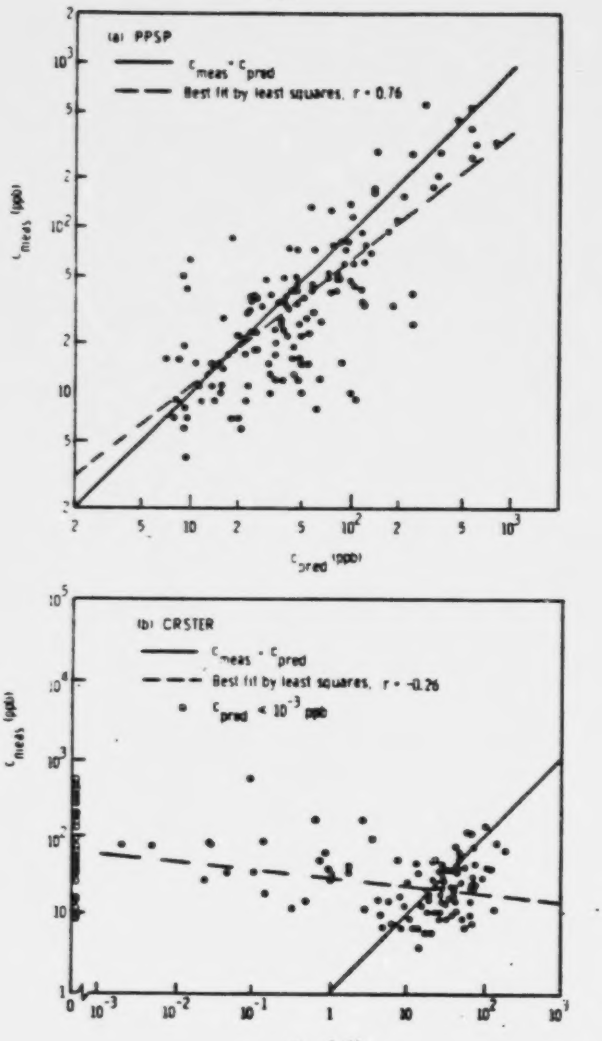


FIGURE 2-1. Comparison between measured and predicted ground-level SO_2 concentrations for the (a) PPSP and (b) CRSTER models. Least squares and correlations are estimated using $\log_{10}C$. (Source: Weil and Brower, 1984)

Hanford 67 (Idaho)
Rancho Seco 4 (California)
Karlsruhe (Germany)
Goodyear (Ohio)
Benicia (California)
Garfield (Utah)

The models that were evaluated included four rural models (PTMTP, PAL, CRSTER, RAMR) and one urban model (RAM). Standard statistics were computed, including the following:

Ratio of mean predictions over mean observations

Correlation coefficients (matched in space or time)

Ratio of average of 10 highest predictions over average of 10 highest observations

Percent bounded relative errors within the three ranges < 0.2, 0.5-2.0, and > 5.0.

The performance results were stratified by unstable, neutral, and stable meteorological conditions. The averaging period for the concentrations ranged from 10 minutes to one hour.

The P/O ratios for the 10 highest observations and predictions were of most interest to our study. The model performance statistics for the elevated releases are summarized in Table 2-5. From this table we can see that there is substantial variation in the P/O ratios among the tracer experiments, with the Hanford and Benicia experiments representing the extremes of overprediction and underprediction respectively.

The models tend to overpredict for unstable conditions and underpredict for stable conditions. Differences in performance between the models are not significant, with the single exception that the urban version of RAM has more of a tendency to overpredict than do the other rural models, particularly under neutral and stable conditions.

The only two point source evaluation studies that involved both elevated release and buoyant emissions were at Goodyear and Benicia. All of the rural models substantially overpredicted the Goodyear observations under unstable conditions, but underpredicted when conditions were neutral or stable. The urban model RAM overpredicted for all stability categories. All of the rural models substantially underpredicted

TABLE 2-5. The ratios of the average 10 highest predictions and average 10 highest observations from the API model performance evaluations using observations from elevated tracer releases.

Model/ Condition*		Hanford 67	Rancho Seco	Karlsruhe	Goodyear	Benicia	Garfield
PTMTP	U	2.61	2.03	0.85	2.05	0.57	0.07
	N	2.44	0.86	0.18	0.20	0.14	-
	S	1.56	0.65	-	0.21	0.23	-
PAL	U	2.26	1.63	0.57	2.19	0.44	0.14
	N	1.88	0.76	0.10	0.40	0.34	-
	S	0.90	0.51	-	0.38	0.18	-
CRSTER	U	2.37	1.84	0.54	2.22	0.41	0.86
	N	1.82	0.88	0.09	0.44	0.36	-
	S	1.18	0.76	-	0.38	0.19	-
RAMR	U	2.45	1.81	0.70	2.16	0.55	0.11
	N	2.09	0.80	0.13	0.34	0.27	-
	S	0.94	0.52	-	0.32	0.16	-
RAM	U	1.30	2.46	0.90	2.25	0.45	0.20
	N	0.71	2.23	0.65	1.62	0.45	-
	S	1.07	2.01	-	2.69	1.16	-

* U = unstable, S = stable, N = neutral

the Benicia data for all stability categories, whereas the urban model RAM slightly overpredicted for stable conditions.

EPRI Plume Model Validation and Development Study

The Electric Power Research Institute's Plume Model Validation and Development (PMV&D) Study collected field data at two sites. The first experiment was in flat terrain at Kincaid, Illinois. The validation study was conducted using SF_6 tracer release data from an intensive study and SO_2 collected during routine monitoring by a network of 28 fixed monitors. Three models, CRSTER/MULTIMAX (CEQM), TEM, and MSDM, were evaluated using versions supplied by developers. Three types of meteorological inputs were used: (1) NWS off-site meteorology, (2) routine on-site meteorology, and (3) best estimates of on-site meteorology. We focused on the best estimates of on-site meteorological data required for the CRSTER/RAMMET preprocessor.

We are primarily interested in the highest unpaired data and the time-matched 1-hour data. Table 2-6 summarizes the results for the performance evaluations using unpaired hourly averaged SO_2 data from Kincaid. CEQM slightly overpredicts the highest and second highest concentrations whereas TEM underpredicts. Both models underpredict the average of the highest 25 concentrations. However, CEQM appears to be very close (within 5 percent) using the input data set known as level 1b. CEQM tends to underpredict for high wind conditions and overpredict for low wind speed conditions, when the highest concentrations are observed. When the data are stratified by stability class (Figure 2-2), the CEQM strongly underpredicts during neutral and stable conditions. The TEM underpredicts, but the underprediction is not as strongly dependent on stability class as that of CEQM. Overpredictions occur close in to the source (2 km or less), and underpredictions occur beyond 5 km. The CEQM shows an even more pronounced tendency to overpredict the highest concentrations of SF_6, while TEM underpredicts these. The MSDM model strongly overpredicts the highest values of both SF_6 and SO_2 by more than a factor of two.

The time-matched 1-hour average tracer predictions made by both CEQM and TEM are significantly less than observations; less than half the predictions are within a factor of two of the observations. The correlation coefficients are poor; none are greater than 0.30. The average absolute error is about 80 percent of the average observed concentration. The results are similar for SO_2 except that the correlations are even smaller, and the average absolute error is roughly the same as the average observed concentration.

The second model evaluation program of the PMV&D study used field data collected at a power plant in Bull Run, Tennessee, located in rolling terrain. The winds were substantially lower than those at Kincaid. There was no separate model performance evaluation using SO_2; only the SF_6 tracer data was used. Except for CEQM, the models evaluated were not those evaluated with the Kincaid data.

TABLE 2-6. Comparison of 25 highest 1-hour average observed and predicted SO_2 concentration values (ppb) at Kincaid, Illinois, that are unpaired in time and location. (Source: Reynolds et al., 1984).

Observations/ Predictions	Input Level	Single Highest Value	2nd Highest Value	7th Highest Value	13th Highest Value	19th Highest Value	25th Highest Value	Average of 25 Highest Values
Observed value		459	450	333	273	249	231	301
Predicted value								
CEQM	Ia*	459	387	250	213	198	187	237
CEQM	Ib	602	533	299	254	230	220	291
CEQM	II	589	415	303	263	239	205	285
TEM	Ia	308	285	241	213	197	184	221
TEM	Ib	330	312	274	263	235	220	260
TEM	II	303	294	250	217	204	196	231
MSDM	Ia	1974	1485	1158	840	682	635	943
MSDM	Ib	6449	3608	2009	1568	1333	1203	1909
MSDM	II	4908	4195	2125	1618	1015	910	1898

* Type of meteorological input used: Ia is a NWS data source, Ib is an onsite data source, II is the best method of estimating needed input data from available observations.

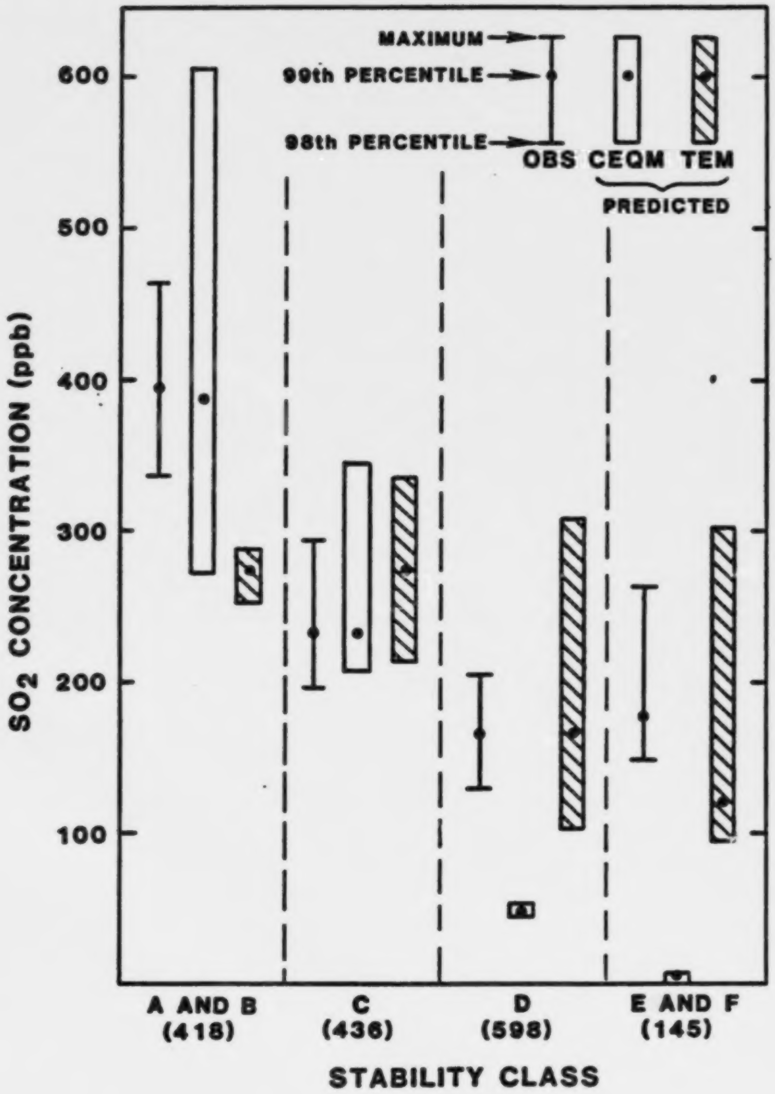


FIGURE 2-2. Maximum 1-hour SO_2 concentration for each hour categorized by stability group. The data set is for hours at Kincaid, Ill., when SO_2 concentrations were predicted. (Source: Reynolds et al., 1984)

Figure 2-3 summarizes the findings for the highest unpaired data sets. From this figure we can see that all models underpredict the highest observed concentration. All models except CEQM underpredict the second highest concentration, and all models except RTDM underpredict the fifth highest observed concentration. When the unpaired data are stratified by stability class we find that

> CEQM slightly overpredicts during class A conditions but strongly underpredicts under class B, C, and D conditions.
>
> SHORTZ always underpredicts except during class E and F conditions, when it produces a suspiciously large prediction.
>
> RTDM does not significantly over- or underpredict the highest concentrations except under class D conditions, when it severely underpredicts.

When observations and predictions are paired in time, all three models tend to underpredict. The average absolute residual is slightly more than half the observed average concentration (Table 2-7). The correlation coefficients are significantly larger than those found at the Kincaid site. RTDM predictions have the greatest correlation with observations (0.70), and over 40 percent of its predictions are within a factor of two of the observations. The SHORTZ model predictions seem to have the greatest imprecision and the least correlation with the observations.

EPA Rural Model Evaluation Study

The EPA rural model evaluation study is an ongoing program. To date, data from power plant sites at Clifty Creek, Kincaid, Muskingum River, and Paradise in the Midwest have been used to evaluate model performance. The first model evaluation study used 1975-1976 data from Clifty Creek and evaluated 10 models (see Table 2-3). In later studies under this program the number of models evaluated was pared to four (PPSP, MPTER, MPSDM, and TEM-8A) based on recommendations by Smith (1984). The evaluation statistics were based on those recommended by the AMS workshops on model evaluation (Fox, 1981). A statistical packaged developed by TRC in the first EPA rural study has been used throughout the program.

The rural model evaluation study has produce volumes of statistics, which by themselves are hard to review and synthesize. The statistical package is automatic, and as a result some of the stratified statistics make little sense. For example, the modeling results were stratified by wind speed and stability class; however, the stratification was done for data other than 1-hour averaged data and for unpaired data. A 3- or 24-hour stability class is relatively meaningless, as is the comparison between meteorological condition and unpaired statistics. In our review we focus on the N highest predictions unmatched to observations and on the highest pairs of predictions and observations matched in time and stratified by meteorology.

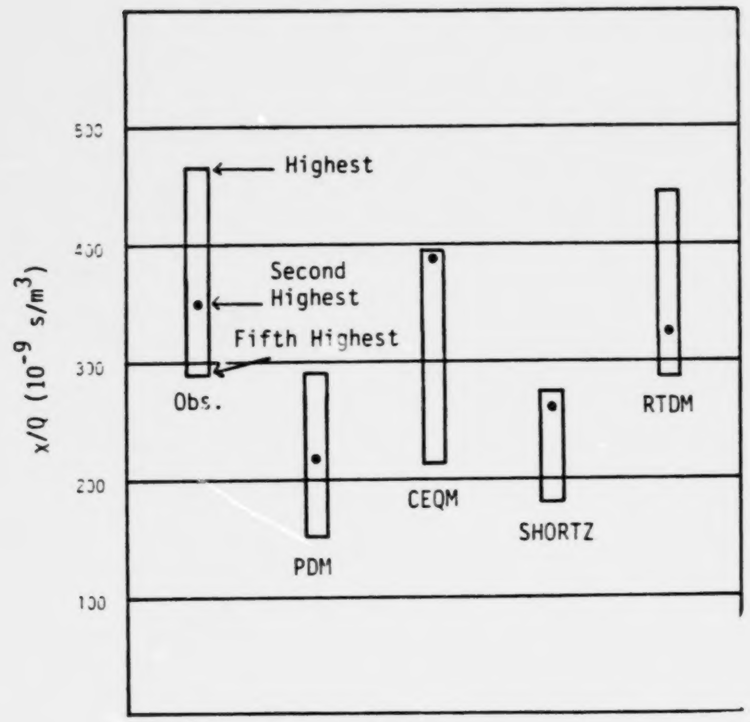


FIGURE 2-3. Five highest 1-hour average observed and predicted χ/Q values (10^{-9} s/m^3) unpaired in time or location--all data in the common data set (93) hours). Results are for the Bull Run, Tennessee site located in moderately complex terrain. (Source: Reynolds et al., 1985)

TABLE 2-7. Highest 1-hour average observed and predicted x/Q (10^{-9} s/m^3) paired in time--all data in common data set. Results are from the Bull Run, Tennessee site located in moderately complex terrain. (Source: Reynolds et al., 1985)

			Performance Measurement				
Model	Number of Pairs Above Threshold	Average Observed Value	Average Residual* (Obs.-Pred.)	Median Residual (Obs.-Pred.)	Average Absolute Residual	Correlation Coefficient*	Percentage of Predictions Within a Factor of Two of the Observations
PDM	93	100	38 ± 18	18	58	0.53 (0.37; 0.67)	48
CEQM	93	100	35 ± 21	20	63	0.54 (0.38; 0.67)	40
SHORTZ	93	100	64 ± 18	48	82	0.16 (-0.05; 0.35)	19
RTDM	93	100	37 ± 22	27	59	0.70 (0.58; 0.79)	41

---

* Listed with the 95 percent confidence interval.

The original evaluation, based on the 1975-1976 Clifty Creek data, found that for hourly SO_2 concentrations:

> The correlations for predicted and observed concentrations matched by time and location were low (not significant from zero).
>
> All of the models performed roughly the same.
>
> The models tended to overpredict for unstable conditions and underpredict for stable conditions.
>
> On a model-by-model comparison, COMPTER seems to have the smallest fractional difference and the smallest standard deviation in the residuals. This difference, however, was not outstanding.
>
> For neutral and stable conditions, all models except COMPTER underpredicted.
>
> MPSDM performance was variable; it underpredicted for some sites and came close for others.
>
> The 3141/4141 model always underpredicted regardless of wind speed and stability.

While unpaired data can in principle be stratified by wind speed and stability class, the interpretation of the results must keep in mind that, since there is no matching by time, other data inputs, such as mixing height, can confuse attempts to evaluate model performance. Since there are no data on the highest concentrations matched by time, we examined the unpaired highest 25 concentrations stratified by wind speed and PGT stability class. The results (Table 2-8) indicate the following general behavior:

> There is general consistency between evaluation statistics for the 1975 and 1976 data.
>
> All models except 1314/1414 either overpredict or show no significant bias.
>
> The models overpredict for the most unstable conditions and underpredict for the most stable conditions.
>
> The PPSP model produces excessive overprediction for all stability classes and wind speed ranges.
>
> The greatest underprediction by all of the models occurs when wind speed is high.

TABLE 2-8. Average residuals (observations minus predictions) of the 25 highest observed 1-hour SO_2 concentrations ($\mu g/m^3$) for the performance evaluation of models using the 1975-1976 Clifty Creek data sets. (Compiled for this study from Londergan et al., 1982.)

			Wind Speed Stratification		PGT Stability Class Stratification	
Model	(Year)	All	Maximum	Minimum	Maximum	Minimum
MPTER	(75)	-207 (-403, -170)	-68 (H)	-369 (M)	56 (EF)	-481 (AB)
	(76)	-313 (-369, -257)	-161 (H)	-285 (M)	155 (D)	-462 (AB)
MPSDM	(75)	-423 (-643, -204)	167 (H)	-517 (L)	361 (EF)	-552 (AB)
	(76)	-233 (-326, -140)	138 (H)	-331 (L)	323 (EF)	-329 (AB)
COMPTER	(75)	-339 (-455, -224)	-129 (H)	-420 (M)	-32 (D)	-522 (AB)
	(76)	-375 (-428, -322)	-183 (H)	-380 (M)	103 (D)	-511 (AB)
SCSTER	(75)	-69 (-186, 47)	259 (H)	-112 (M)	350 (D)	-358 (AB)
	(76)	-135 (-205, -65)	140 (H)	-157 (L)	422 (D)	-352 (AB)
3141/4141	(75)	234 (131, 336)	308 (H)	139 (L)	415 (D)	-61 (AB)
	(76)	206 (152, 260)	259 (H)	146 (L)	522 (D)	-17 (AB)
TEM-8A	(75)	-161 (-310, -12)	275 (H)	-133 (L)	389 (EF)	-193 (AB)
	(76)	39 (-46, 121)	306 (H)	-42 (L)	454 (D)	-19 (AB)
MULTIMAX	(75)	-49 (168, 70)	218 (H)	-83 (L)	313 (D)	-338 (AB)
	(76)	-120 (-193, -46)	136 (H)	-154 (L)	400 (D)	-348 (AB)
PPSP	(75)	-3210 (-3500, -2930)	-114 (H)	-3341 (L)	-581 (EF)	-2971 (D)
	(76)	-3571 (-4005, 3137)	-184 (H)	-3674 (L)	-850 (EF)	-2628 (D)

(1975) $\bar{X}_{obs} = 772$
(1976) $\bar{X}_{obs} = 790$

Key:

L = $W_s < 2.5$ m/s
M =$2.5 < W_s < 5.0$ m/s
H =$W_s > 5$ m/s

Notes:

- = overprediction
+ = underprediction
(,) = low, high 95 percent confidence interval

> The greatest overprediction generally occurs when wind speed is low (unstable conditions probably dominate this category) because predicted and observed concentrations are larger under low wind speed.

The rural model study was continued using data from Muskingum River, Kincaid, and Paradise. We examined the performance statistics for the top 25 concentrations from these studies hoping to find some systematic results. The performance statistics for these evaluations (Table 2-9) indicate that PPSP grossly overpredicts all observed data sets. The MPSDM model also as a rule strongly overpredicts. MPTER overpredicts the Muskingum River data, but shows no significant bias against the Kincaid and Paradise data sets. The TEM-8A model generally produces the smallest bias of the four models.

The statistics in Table 2-9 further demonstrate that maximum overpredictions occur under unstable conditions and maximum underpredictions under neutral or stable conditions. However, while for some sites the maximum overprediction occurs during highly unstable conditions, for others it occurs under class C and D conditions, or even E and F (in the case of PPSP). The maximum overpredictions by the MPSDM and PPSP models occur when wind speed is low, whereas those by MPTER and TEM occur when wind speed is moderate or high.

EPA Urban Model Evaluation Study

The urban model evaluation study was conducted using 1976 RAPS data for St. Louis. The SO_2 air quality data for the validation comes from 13 monitoring stations. Two models, TEM-8A and RAM (UNAMAP 3 versions), were evaluated for predictions of 1-hour SO_2 concentrations.

The performance statistics for the 25 highest unpaired predicted and observed concentrations (Table 2-10) indicate that, overall, TEM overpredicts while RAM underpredicts. The variance of the RAM predictions is more realistic than that of TEM-8A. For TEM-8A the RMSE of the highest concentrations at each site is 2295 $\mu g/m^3$, while for RAM it is less than half that (993 $\mu g/m^3$). The correlation coefficient for the highest at each site is zero for TEM-8A and negative for RAM. The RAM produces smaller residuals than TEM-8A for all wind speed and stability class ranges examined. The greatest overpredictions (or smallest underpredictions) occur under low wind speed and stable conditions. The largest underpredictions (smallest overpredictions) occur during high wind speed and neutral or slightly unstable conditions.

Other analyses have examined specific aspects of the RAM model performance. For example, Ruff (1983) regressed residuals (observations minus predictions) versus model input parameters from RAM applications to several sites in an attempt to explain the residuals. One of Ruff's findings was that atmospheric stability seemed to be a major source of variance in the residuals. Ruff also suggested that the lumping of nearby emission sources could be another significant source of error. In a

TABLE 2-9. Average residuals (observations minus predictions) of the 25 highest observed and predicted unpaired 1-hour SO_2 concentrations ($\mu g/m^3$) for selected data bases. (Compiled for this study from Cox and Moss, 1985; Cox et al., 1986, 1987).

		Wind Speed Stratificatgion		PST Stability Class Stratification	
Model	All	Maximum	Minimum	Maximum	Minimum
Muskingum $\bar{x}_{obs}$ = 624) (1975 data)					
MPTER	-464 (-543, -384)	-113 (H)	-530 (M)	118 (D)	-575 (AB)
MPSDM	-341 (-404, -277)	50 (H)	-374 (M)	113 (EF)	-402 (AB)
PPSP	-1352 (-1498, -1205)	-509 (H)	-1165 (L)	-293 (EF)	-1245 (D)
TEM-8A	-57 (-137, 23)	225 (H)	-150 (M)	361 (D)	-126 (C)
Kincaid ($\bar{x}_{obs}$ = 760) (1980-1981 data)					
MPTER	37 (-74, 148)	213 (L)	-64 (M)	250 (EF)	-34 (C)
MPSDM	-418 (-505, -331)	53 (H)	-470 (L)	255 (EF)	-500 (C)
PPSP	-3361 (-3700, -3021)	-603 (H)	-3345 (M)	-1680 (D)	-3235 (EF)
TEM-8A	76 (6, 147)	97 (L)	-21 (M)	203 (EF)	-58 (C)
Paradise ($\bar{x}_{obs}$ = 1590) (1976 data)					
MPTER	-149 (-336, 38)	266 (L)	-284 (H)	1008 (D)	-319 (AB)
MPSDM	-1309 (-1572, -1046)	-357 (H)	-1484 (L)	269 (EF)	-1222 (AB)
PPSP	-3889 (-4601, -3176)	-1191 (H)	-3989 (L)	-479 (EF)	-3342 (AB)
TEM-8A	85 (-127, 296)	332 (M)	-70 (H)	671 (D)	-10 (C)

Key:

L = $w_s < 2.5$ m/s
M = $2.5 < w_s < 5.0$ m/s
H = $w_s > 5$ m/s

Notes:

$-$ = overprediction
$+$ = underprediction
(,) = low, high, 95 percent confidence interval

TABLE 2-10. Performance statistics from the EPA urban model evaluation study for the highest 25 unpaired 1-hour SO_2 concentration observations and predictions. (Compiled for this study from Londergan et al., 1983).

Statistic	TEM-8A	RAM
Average observed	1929	1929
Average residual (obs. minus pred.)	-2069 (-2460, -1678)	307 (137, 477)
Variance comparison	0.07	0.47
Wind speed stratification		
Maximum residual	-96 (H)	409 (H)
Minimum residual	-2327 (L)	51 (L)
Maximum ratio	3.9 (H)	4.9 (H)
Minimum ratio	0.1 (L)	0.7 (L)
Stability class stratification		
Maximum residual	202 (C)	821 (D)
Minimum residual	-2196 (EF)	152 (EF)
Maximum ratio	8.4 (AB)	22.2 (AB)
Minimum ratio	0.9 (EF)	0.7 (EF)

Key

$H = w_s > 5.0$ m/s

$M = 2.5 < w_s < 5.0$ m/s

$L = w_s < 2.5$ m/s

review of the RAM model performance for St. Louis, Turner et al. (1985) point out that most of the highest predicted 1-hour SO_2 concentrations fall within a factor of two of the observations.

Other Model Performance Studies

Of the miscellaneous model studies listed in Table 2-2 we found only one, the Michigan Downwash Study, with the site characteristics we desired for our study. In the Michigan study the unmodified RAMR model was found to underpredict the observed concentrations by about 50 percent. When downwash was added to the model, the model performance became irregular, overpredicting for some monitors and underpredicting for others. The model slightly underpredicted the highest and second highest concentrations.

The study site, the Widow's Creek plant, is located southeast of an area of complex terrain. When the performance of MULTIMAX was compared to data collected at this site, the model could not predict high concentrations under plume impingement (Lott, 1984). False high concentrations were predicted during stable conditions because the plume rise algorithm used an unrepresentative temperature lapse rate.

The VISTTA study (White et al., 1985) essentially evaluated visibility parameters, although some attention was paid to evaluation of crosswind integrated concentrations (CWIC) of NO_2. Model-predicted and observed CWICs generally had correlation coefficients of greater than 0.5.

A performance evaluation study of ISC and BLP using data collected at an aluminum reduction plant (Scire and Shulman, 1981) indicates that BLP overpredicts the highest 1-hour SO_2 concentrations, with fractional differences of approximately -0.25. The ISC model strongly underpredicts, with fractional differences of approximately 1.0. A comparison of the 10 highest unpaired 1-hour concentration averages for BLP indicates that BLP overpredicts, with a fractional difference of the order -0.4 to -0.6. The correlations for the 50 highest unpaired observed and predicted concentrations are of the order 0.7 to 0.9.

In an evaluation of ISC, MPTER, and BLP, also using data from an aluminum plant site (McTaggert, 1981), the five highest predictions of each model are similar. The performance of these models is variable; concentrations at some sites are overpredicted, with maximum fractional differences as large as -0.9.

Of all of the models evaluated in limited studies, the RAM model has been most frequently evaluated (Table 2-11). Although most of these studies were done for a large elevated point source in an urban setting, the EPA urban evaluation study at St. Louis was done in an area where no single point source dominates the air quality at all the stations. Several studies show that RAM overpredicts when it is used to model a major elevated and buoyant point source in an urban environment. This is in

TABLE 2-11. A summary of RAM performance evaluation studies.

Location/Reference	Major Findings
St. Louis R. E. Ruff, H. S. Javitz, and J. S. Irwin (1980)	(hourly data) - Matched location; time correlations at stations ranged from 0.10 to 0.23 - RAM overpredicted and overpredicted more with increasing stability
Cleveland R. F. Hodabosi and R. Brodkey (1984)	(24 hour data) - Inconclusive; overpredictions occur on some days, and underpredictions on others; no reasons were identified
Cleveland R. F. Hodabosi and L. K. Peters (1981)	(24 hour data) - Station correlations of time-matched predictions; observation pairs ranged from -0.121 to 0.541 - Second-highest predictions were as great as 180% of the observed or were underpredicted by as much as 27% of the observed
East Lake and Avon Lake Plants near Cleveland H. M. Ellis and P. C. Liu (1981)	(hourly data) - Overpredictions by a factor of 1.7-2.2 at sites around East Lake Plant - Overpredictions by a factor of 1.2-2.1 at sites around Avon Lake Plant
Michigan City, Indianapolis P. H. Guldberg and C. W. Kern (1978)	(24 hour data) - RAM overpredicted the maximum concentrations by approximately a factor 1.8 and had an average overprediction of 105% at Michigan City - RAM overpredicted maximum concentrations by an excess of a factor of 4 at Indianapolis
Detroit R. H. Kummler, B. Cho, G. Roginski, R. Sinha, and A. Greenberg (1979) Detroit	(3 hour data) - Correlation of time- and location-matched concentrations was 0.42 - The RAM produced maximum concentrations close to those that were observed

contrast to the findings of the EPA urban evaluation study, which showed RAM slightly underpredicting.

SELECTION OF STUDIES FOR FURTHER REVIEW

Our review of model performance evaluations turned up a number of data sets that would lend themselves to statistical analysis of concentration residuals as a function of meteorology. After ruling out data for sites in complex terrain or near interfering point sources (so that background estimation becomes a challenge), we ended up with the list of data sets in Table 2-12. Some of the data sets that we turned down and the reasons for their rejection are listed in Table 2-13. Table 2-14 summarizes the amount of data in each data set. Once the list of studies for futher review had been agreed upon, we attempted to acquire complete data sets that include the meteorology, observed air quality, and predicted concentration data.

The original owners of the data sets were contacted. Some of the data sets listed were no longer available, e.g., the Michigan Power Plant study data. The API Tracer Study data tape received by Systems Applications from TRC contained only observations of meteorology and air quality, but no predictions from any of the models evaluated by Londergan and others (1980). Apparently the model predictions are no longer available for use. When we attempted to acquire the high wind data for the Baldwin plant, which had been collected and analyzed by ERT, we were informed by ERT that the data set is still confidential and would not be available for this study.

As a result, the data sets available for analysis shrank to the following four:

> The EPRI Plume Model Validation and Development data sets for Kincaid, Illinois and Bull Run, Tennessee
>
> The Clifty Creek 1975 and 1976 data sets used by Mills and others (1979)
>
> The Maryland Power Plant Siting Program data set described by Weil and Brower (1982)

These data sets were used to evaluate the following models:

Data	Models
Kincaid	CRSTER Equivalent Multimax (CEQM), TEM, HPDM
Bull Run	CEQM, SHORTZ, RTDM, HPDM
Clifty Creek	MPTER/CRSTER
Maryland	PPSP--CRSTER, PPSP

In the next section we review the formulation of these models.

TABLE 2-12. Data sets selected for further study.

Location (Sponsor)	Parameters*	Models	Reference
Kincaid, Illinois April-July 1980 routine 20 April-10 May, intensive 9 July-29 July, intensive	All, hourly; routine/intensive; 30 stations, met tower, upper air data for 150 hours; stability, turbulence	CRSTER TEM CEQM MSDM	Reynolds et al., 1984
Dickerson, MD Chalk Point, MD Morgantown, MD 1973-1974 (Martin-Marietta)	All, hourly; for tall stacks (>100 m); surface observations and rawinsondes from Washington National and Dulles International airports	CRSTER PPSP	Weil and Brower, 1982
Goodyear, OH Benicia, CA (API/TRC, Inc.)	All, hourly; tracer experiment; elevated, buoyant sources	CRSTER; PAL, RAMR; PTMPT; RAM	Londergan et al., 1980
Bull Run, Tennessee 28 July-24 August 1982 22 September-19 October 1982 (EPRI PMVD)	All, hourly; tracer	CEQM: RTDM SHORTZ; SAI-PDM	Reynolds et al., 1984
Michigan Power Plant 1978-1979	All, hourly	RAMR (modified)	Touma et al., 1981
Clifty Creek Plant, IL 1975-1976	All, hourly	CRSTER; MPTER; 3141/4141; SCSTER; PLUMES; MPSDM, TEM8A, COMPTER; MULTIMAX	Londergan et al., 1982

* "All" means all data needed to exercise the model and evaluate its performance.

TABLE 2-13. Data sets rejected for further study.

Data Set	Reasons for Rejection	Models
Westvaco-Luke Mill, WV December 1979 - November 1981 (HE Cramer Co.)	Complex terrain	VALLEY, COMPLEX I, II; SHORTZ; LUMM
Westvaco-Luke Mill, WV December 1980 - November 1981 (TRC, Inc.)	Complex terrain	COMPLEX I, II; PFM; 4141; PLUME5; RTDM; SHORTZ; IMPACT
Cinder Cone Butte, ID 16 October 1980 - 11 December 1980 (ERT)	Complex terrain	COMPLEX I, II; PFM; 4141; PLUME5; RTDM; SHORTZ; IMPACT
Garfield, UT (API/TRC, Inc.)	Complex terrain	CRSTER; RAM; RAMR; PAL; PTMPT
Green Glow; NRTS; Prairie Grass; TMI; Shoreline; St. Louis; Paramount; Rancho Seco; Karlsruhe (API/TRC, Inc.)	Non-elevated and/or non-buoyant sources	CRSTER; RAM; RAMR; PAL; PTMPT
Philo, Muskingum, OH 1 January 1973 - 31 December 1973 (GCA Technical)	Plants located within 40 km of one another	CRSTER
Ohio Edison SAMMIS Plant (Enviroplan)	100 point sources within 30 km	58 models evaluated

Continued

TABLE 2-13 (concluded).

Data Set	Reasons for Rejection	Models
JM Stuart Plant, OH 1 January 1973 31 December 1973 (GCA Technical)	Meteorological data from distant sources (90 km and 130 km)	CRSTER
Widow's Creek, TN July 1978 - February 1979 (TVA)	Plant is located in complex and elevated terrain	TVA-QSM MULTIMAX (modified)
St. Louis RAPS study 1976 (SRI International)	Plant is located in an urban area	RAM

TABLE 2-14. A summary of the meteorological data reported during the model evaluation exercises selected for further study.

Data Set	Meteorology
Kincaid, IL (EPRI PMV&D)	
10 m tower:	T (2-10)
100 m tower:	W_s, W_d (10, 30, 50, 100) T (10-50, 10-100) uvw (100) T_d (100)
On-site:	T (z), W_s (z), W_d (z)
Sodar:	Mixing height, uvw
Dickerson, Chalk Point, Morgan Town, MD (Martin-Marietta)	
On-site:	W_s (z), W_d (z) Mixing height assumed = lower bound of dT/dz isothermal above dT/dz < 0 for at least 50 m
Off-site:	W_d, W_s from Washington National and Dulles International airports; stability (via PGT methods) calculated from these data
API/TRC Inc.	
Location unknown	W_d, W_s, stability, temperature, pressure, mixing height; W_d derived from χ_{max} pattern
Bull Run, TN (EPRI PMV)	
10 m tower:	T (2, 10); uvw (10)
122 m tower:	W_s, W_d (10, 30, 100, 122) T (10-50, 10-100, 30-122) T_d (100), uvw (100)
On-site:	Mixing height (2)
Off-site:	W_s, W_d (5 NWS locations) W_s, W_d (7 external sites)
Michigan Power Plant	
90 m tower:	T (10, 60, 90)
On-site:	Mixing height; W_s, W_d; T
Clifty Creek, IL	
60 m tower:	T (10), W_s (7 m scaled from 60), W_d (60)
On-site:	Mixing height from Dayton NWS site; cloud cover, height from Cincinnati NWS site

3 REVIEW OF MODEL FORMULATIONS

INTRODUCTION

In this section we review how the models included in our study are formulated to simulate ground-level concentrations under the meteorological conditions identified as leading to maximum ground level concentrations. The meteorological conditions that were identified included:

(a) Calm or very low wind speeds during convective conditions;
(b) Moderate to high wind speeds during convective conditions;
(c) Fumigation or low mixing heights during convective conditions;
(d) High wind speeds during neutral or slightly stable conditions.

It should be noted that models like CRSTER have undergone a continuous process of revision. The formulations presented here are those used during the model evaluation studies. Each model also has a number of options that can be exercised, and it is not always clear from the published results of evaluation studies which options were used for a study. We will note wherever possible when options or features that are not recommended in the U.S. EPA Guidelines on models were used in a model evaluation study.

The models reviewed, and the data sets used in the performance evaluations of these models, are listed in Table 3-1.

As noted above, meteorological conditions that are of primary interest are those associated with convective conditions. As such, the assumptions made about the structure of the atmosphere in the vertical can have a substantial influence on the predicted field of concentrations. This structure is defined by assumptions made about wind, temperature, and turbulence profiles and the depth of the mixed layer. Observations can be used to specify these quantities, but widespread application of these models requires that a reasonable default structure be prescribed when only rudimentary, near-surface data are available. The defaults vary somewhat among the models and therefore influence model performance. The extent to which a model uses observed mixing heights, temperatures, etc., rather than default values varies with the data base.

TABLE 3-1. Air quality models and data sets included in model evaluation studies.

Data Set	Models
Kincaid	CRSTER (CEQM)*, TEM, HPDM
Maryland	MPTER, PPSP
Bull Run	CEQM, RTDM, SHORTZ, HPDM
Clifty Creek	CRSTER

* CEQM is a CRSTER-equivalent version of MULTIMAX. CEQM may be viewed as a multiple-source version of CRSTER that has the ability to accept any user-specified receptor arrays rather than the fixed polar grid available in CRSTER.

In addition to different assumptions about the vertical structure of the atmosphere, the models also differ in how the plume rises and disperses in the vertical. Consequently, formulations of how plume material is able to rise through or penetrate the top of the mixed layer vary as well. The manner in which spreading of the plume (horizontally as well as vertically) is treated also varies among the models.

Not all of the data sets are for sites located in areas of flat terrain. The Bull Run site includes significant terrain features, although none of these exceed plume height. Consequently, the manner in which the effect of terrain is incorporated in the models must also be considered.

We first review the structure of the atmosphere in the vertical in each model. We then review the response of a plume in this atmosphere, including plume rise and dispersion. This is followed by a description of the manner in which terrain is included in the models. In each case CRSTER (CEQM) is discussed first so that differences between this "benchmark" model and others can be identified. Finally, the treatment (explicit or implicit) of "calm" situations is discussed.

The major components of the models are listed in Table 3-2 along with each of the four meteorological categories identified. Although each of the components affects model performance in each of the categories, the X's in the table identify those components that are particularly important for a category.

TABLE 3-2. Components particularly important to certain meteorological categories.

Component	Low Wind Speed Convective	Moderate Wind Speed Convective	Low Mixing Height Convective	High Wind Speed Neutral
Plume Rise	X	X	X	
Partial Penetration	X		X	
Meteorological Profiles		X		X
Sigma-z	X		X	X
Sigma-y				X
VDF (pdf versus Gaussian)	X		X	

VERTICAL STRUCTURE

CRSTER (CEQM). CRSTER (EPA, 1977) contains, at most, two layers in the vertical, separated by the mixing height. Dispersion calculations are performed only for plumes whose centerlines are in the lower (mixed) layer (after reaching final rise). This lower layer is characterized by a single wind speed, wind direction, and stability class. The wind speed used in the layer is that which is either measured at or extrapolated to stack-top. The extrapolation method assumes a power-law profile so that

$$u(z) = u(z_m)\,(z/z_m)^p$$

where z_m is the height at which wind speed is measured, and p takes on a value that depends upon the stability class:

A: 0.10 B: 0.15 C: 0.20 D: 0.25 E: 0.30 F: 0.30

These coefficients are historical values used in the model evaluation studies, but have been changed in the UNAMAP6 version.

The mixing height is treated as a reflecting surface for plume material if the stability is neutral or unstable (classes A-D). In stable conditions, no limit is placed on the dispersion of the plume in the vertical. The reflection of plume material from the mixing lid is simulated by the image source method. Multiple reflections are computed between the mixing lid and the surface until the vertical spread of the plume equals 1.5 L, where L is the mixed-layer depth, i.e., the mixing height. Beyond this point, the plume is uniformly mixed in the layer.

TEM. TEM (TACB, 1979) assumes the same layered structure as CRSTER. However, the determination of which plumes exit the mixed layer is different. In CRSTER the plume is placed above the mixed layer if the final height of its centerline exceeds the mixing height. TEM defines a penetration factor I, which is set to 2.0 as a default. If the plume (after rise) exceeds the product IL (i.e., twice the depth of the mixed layer for the default value), the plume is placed outside the mixed layer, and ground-level concentrations are zero. For plume heights between L and IL, the plume height is reset to L. (This difference from CRSTER can be eliminated by setting I = 1.0).

TEM also differs from CRSTER in the distance at which the plume is assumed to be mixed uniformly within the mixed layer. The region of uniform mixing begins at twice the distance to where $\sigma_z = 0.47$ L.

MPTER. The vertical structure of the atmosphere in MPTER (Pierce and Turner, 1980) is very similar to that in CRSTER. The only difference is in the specification of the wind speed profile. The default values of exponents for the power-law formulations for each stability class are:

A: 0.07 B: 0.07 C: 0.10 D: 0.15 E: 0.35 F: 0.35

RTDM. In the Bull Run applications, the RTDM (ER&T, 1982) differs from CRSTER in modeling the vertical structure of the atmosphere. The same two-layer structure is maintained for stabilities A-D, and a single wind speed characterizes plume transport and diffusion. But in stable conditions a critical dividing-streamline height (H_{crit}) is introduced in treating the dispersion around high terrain. Because the terrain around the Bull Run site lies significantly below the elevation of the power plant plume, H_{crit} has only a minor influence on the model comparisons. More will be said about H_{crit} in the section on plume rise and dispersion.

Wind speed profile exponents can be default values or values that the user computes from actual profile data on an hourly basis. The latter choice was made in exercising RTDM using the Bull Run data set. Hourly values of the exponent p are formed by a logarithmic least-squares fitting procedure. Another departure from CRSTER is a restriction on the heights to which the power-law wind speed increase is allowed. In unstable conditions RTDM assumes that the wind speed is constant above one-tenth of the depth of the mixing height. In neutral and stable conditions the height of the maximum wind speed for typical roughness lengths in complex terrain is 200U (10 m), where U (10 m) has units of m/s.

With respect to the mixed layer, RTDM was applied to Bull Run with the option that allows unlimited mixing for stable conditions (classes E and F). For neutral to unstable conditions (classes A-D), RTDM has a partial penetration option that computes the amount of material that lies within the mixed layer.

If the plume centerline lies above the mixing height, the plume height for the material below the mixing height is set to L, the mixing height; otherwise, the plume centerline height remains unchanged. This option was not selected in the Bull Run evaluations. Beneath the mixing height the plume is reflected from the mixing lid several times until the plume becomes uniformly mixed. Although the mathematics of the algorithm is different from that in CRSTER, the results are the same. Uniform mixing is assumed beyond $\sigma_z = 1.37$ L.

SHORTZ. Like CRSTER, SHORTZ (Bjorklund and Bowers, 1982) assumes a two-layer structure in which ground-level concentrations are zero for all plumes that stabilize above the mixing height. However, this applies to stable conditions as well as neutral and unstable conditions. The wind speed at stack-top is used to compute plume rise within the mixed layer, while the wind speed at final plume height is used to compute plume transport. Although the power-law extrapolation method is used in both calculations, unlike CRSTER the exponents can be (and are in the Bull Run evaluation study) obtained from hourly profile data from a tower using a logarithmic least-squares regression equation.

Plume material within the mixed layer is reflected from the surface and the mixing height by means of the image source method. Plume material is uniformly distributed in the layer once the contribution from the image sources at +6L exceed e^{-10}.

PPSP. PPSP (Weil and Brower, 1982) also assumes the two-layer structure used in CRSTER. Differences occur in the computation of wind speed for unstable conditions (classes A-C), and in the allowance for plumes to partially penetrate the top of the mixed layer. The wind speed at stack top for classes D-F is calculated by the same power-law extrapolation procedure used in CRSTER. Also, mixing of the plume in the vertical is unlimited for stable conditions (classes E and F).

The wind speed used in the mixed layer for classes A through C is the value at an elevation equal to one-tenth of the mixing height. This speed is considered the average speed in the layer, and is found by using the power-law extrapolation method with the values of p (the exponent) used in CRSTER. One should note that the stability class and therefore the exponent p is dependent on the layer-averaged wind speed in the model. As a result the computation of the mean speed and the stability requires an iterative solution.

For plumes that "straddle" the mixing height, partial penetration of the top of the mixed layer is permitted by the model. The rise of the plume is computed for a bent-over plume in a stable layer (Briggs, 1984) in which the potential temperature gradient is set to .01 °C/m and the vertical depth of the plume is set equal to the amount of plume rise. The amount of plume material that remains below the mixing height is linearly related to the distance between the mixing height and the lower edge of the plume divided by the full depth of the plume. The estimated elevation of the centerline of the part of the plume that remains in the mixed layer is varied linearly between the height of the plume in the absence of any penetration and the mixing height. The linear weighting is done on the basis of the amount of penetration. For full penetration the plume centerline height is set equal to the mixing height; however, the emission rate is set to zero for this situation in order to simulate plume lofting.

HPDM. HPDM contains the two-layer structure in the vertical for neutral and unstable conditions. For plume material within the mixed layer, the mixing height is treated as a reflecting surface. Plume material above the mixed layer remains aloft. During stable conditions, plumes above the surface layer are allowed to diffuse to the surface.

Wind speeds are extrapolated to plume height by means of the profile equation developed by Businger and co-workers (1971) rather than the power-law method. These equations are used for plume heights no greater than 100 m. At heights in excess of 100 m, the wind speed at 100 m is used.

The vertical gradient of the potential temperature is also extrapolated to plume height. The observed gradient at height 1 is related to the gradient at height 2 by the Stull (1983) scaling formula:

$$\left(\frac{d\theta}{dz}\right)_2 = \left(\frac{d\theta}{dz}\right)_1 \exp\left(-.77\ \frac{Z_2}{H}-\frac{Z_1}{H}\right)$$

where H is the height of the planetary boundary layer. If no observations of the temperature gradient are available, similarity theory is assumed to hold and empirical relations in the form described by Businger and co-workers (1971) are used:

$$\frac{d\theta}{dz} = \frac{\theta^*}{.4z}\ (0.74 - 4.7\ z/L)$$

Partial penetration of the top of the mixed layer is simulated in a manner similar to PPSP. If the calculated plume rise exceeds 0.62 percent of the layer between the top of the stack and the mixing height, then a new plume rise is computed using the stable rise algorithm with a potential temperature gradient of 0.005 °C/m. The formulas for the fraction of material that remains in the layer and the effective height of this material are the same as those in PPSP. For the PDF dispersion module he indirect or 'bumping' plume which impinges at the CBL top is treated by using a distribution of point sources whose strength is adjusted to cancel the upward flux of material from the real sources (e.g., see Hanna et al., 1986).

PLUME RISE AND DISPERSION

All of the field data sets included in this study feature buoyant plumes from tall stacks. Therefore, the characterization of plume rise algorithms presented below is restricted to this class of emissions sources.

CRSTER (CEQM). CRSTER computes final rise for three distinct classes of meteorology. These classes and the associated formulations are:

$\Delta h = 5.0\, F^{1/4}\, S^{-3/8}$, stable-calm (Briggs, 1973)

$\Delta h = 2.6\, (F/uS)^{1/3}$, stable-windy (Briggs, 1975)

$\Delta h = 1.6 F^{1/3}\, (3.5\, x^*)^{2/3}\, u^{-1}$, neutral/unstable (Briggs, 1970)

where

$x^* = 14 F^{5/8}$ for $F < 55\ m^4/s$

$x^* = 34 F^{2/5}$ for $F > 55\ m^4/s$

Transitional plume rise, which is the plume rise at distances from the source less than that needed to achieve final rise, was added to CRSTER as an option in 1979. It is computed as suggested by Briggs (1972) for all three meteorological classes as:

$$\Delta h = 1.6\, F^{1/3}\, x^{2/3}\, u^{-1}$$

The transitional plume rise is used in place of the final rise only if it is less than that at final rise. Computation of stack-tip downwash is also available as an option. The algorithm follows the formulation of Briggs (1973). For wind speeds in excess of w/1.5, where w is the exit velocity, the length 2D (1.5 - w/u) is subtracted from the height of the stack, where D is the diameter of the stack.

The dispersion of the plume is prescribed by Turner's fit to the Pasquill-Gifford (Gifford, 1961) dispersion curves for σ_y and σ_z. For buoyant plumes no allowance is provided for initial plume growth during the plume rise phase.

TEM. Plume rise for buoyant plumes, including both the transitional and final rise, is computed in the same manner as in CRSTER. Stack-tip downwash is also included as an option in a form that is equivalent to that in CRSTER, although the formulation is credited to Gifford (1972).

The dispersion parameters differ from those in CRSTER. A minor difference is the use of a slightly different fit to the Pasquill-Gifford curves (Busse and Zimmerman, 1973). A major difference is the use of a factor to adjust the horizontal dispersion parameter for averaging times longer than 10 minutes. The method follows suggestions from Gifford (1975), who recommends that σ_y be increased by dividing it by a stability-dependent factor, R(s), derived from data provided by Singer (1961):

Stability Class	R(s)
A	0.675
B	0.55
C	0.425
D	0.30
E	0.175
F	0.175

MPTER. In addition to the plume rise and dispersion formulations contained in CRSTER, MPTER includes an option for including the initial growth of a buoyant plume. Following Pasquill (1976), the buoyancy-induced plume size (both σ_z and σ_y) is given by $\Delta h/3.5$. The total plume spread is computed by adding in quadrature the buoyancy-induced spread to the spread obtained from the Pasquill-Gifford curves (i.e., by summing the variances).

RTDM. In RTDM the plume rise algorithms for buoyant plumes, including final rise and transitional rise, are equivalent to those contained in CRSTER, including final rise and transitional rise. The stack-tip downwash algorithm differs in that it contains an adjustment for the change in plume rise resulting from the dilution experienced during the downwash phase as well as the physical change in plume centerline height. The stack-tip downwash option was not selected in the Bull Run modeling.

Plume dispersion can be computed in a number of ways, depending on the options selected. For the Bull Run applications, 22.5° sector-averaging was used to specify the distribution of plume material in the lateral direction, while measured turbulence intensity (i_z) data in the vertical were used in conjunction with the following formulas for σ_z:

$$\sigma_z = i_z\, x\, (1 + .0003\, x)^{-1} \qquad \text{stability E, F}$$

$$\sigma_z = i_z\, x\, (1 + .0015\, x)^{-1/2} \qquad \text{stability D}$$

$$\sigma_z = i_z\, x \qquad \text{stability A, B, C}$$

where x is the downwind distance from the source.

Buoyancy-enhanced plume spread in the vertical is computed as $\Delta h/\sqrt{10}$, and is combined with the spread due to ambient turbulence by summing the variances, as in MPTER.

SHORTZ. Plume rise computations in SHORTZ are based on Briggs's equations, but differ from those in CRSTER. For neutral and unstable conditions distance to final rise is set at 10 stack heights rather than $3.5x_*$. Both approaches have been suggested by Briggs, noting that they give nearly the same performance for most observations of plume rise data. For stable conditions, SHORTZ uses

$$\Delta h = 2.4\, (F/uS)^{1/3}\,, \qquad \text{for } uS^{-1/2} < 10h,$$

$$\Delta h = 1.2\, (F/uS)^{1/3}\left[1 - \cos\left(\frac{10h}{uS^{-1/2}}\right)\right]^{1/3}\,, \qquad \text{for } uS^{-1/2} \geq 10h$$

The first of these expressions is equivalent to the CRSTER equation for the stable-windy meteorological class, except that the factor 2.4 rather than 2.6 is used. The second expression assures that Δh (stable) does not exceed Δh (neutral) as the potential temperature gradient approaches zero.

No allowance is made for transitional plume rise, but stack-tip downwash is included. The plume rise is altered by multiplying by a factor that equals zero if the wind speed equals or exceeds the exit velocity; or unity if the wind speed is less than 1/1.5 times the exit velocity; or D(1 - u/W) otherwise. The SHORTZ user's guide points out that this downwash factor should not be used for situations in which the stack Froude number is less than 1.0, and it may not be applicable for Froude numbers between 1.0 and 3.0.

The dispersion parameters contained in SHORTZ are substantially different from the Pasquill-Gifford curves in CRSTER. Lateral and vertical turbulence intensities are used directly in the formulation rather than coefficients that depend on the stability class. If these quantities are not measured, default values that are a function of the stability class can be used. For the Bull Run applications, observed turbulence data are used.

The vertical dispersion coefficient, σ_z, grows linearly with distance for all distances. For σ_y the distance formulation is more elaborate, giving rise to a $x^{0.9}$ distance dependence well downwind of the source:

$$\sigma_y = i_y \, x_L \left(\frac{x - .1x_L}{.9x_L} \right)^{.9}$$

where x_L is the distance over which the lateral growth of the plume is linear with distance. A value of $x_L = 50$ m is suggested.

Buoyancy effects on the initial size of the plume are incorporated by adding a virtual distance to the actual distance from the source, e.g. ($x = x_{virt.} + x_{act.}$). This added distance allows the calculated plume size to expand further to match the buoyancy-induced plume size at the distance at which the final plume height is attained. For neutral and unstable conditions, final rise occurs at 10 stack heights (10h). For stable conditions, final rise occurs at the minimum distance of either $uS^{-\frac{1}{2}}$ or 10h. At this distance, both σ_y and σ_z equal $\Delta h/4.3$ (rather than $\Delta h/3.5$ as in MPTER).

PPSP. In the PPSP model the method of estimating the final plume rise for buoyant plumes in stable conditions is the same as that in CRSTER. For unstable and neutral conditions the plume rise is taken to be the lower of the predictions of Briggs's (1975) breakup and touchdown formulations. For neutral conditions, the breakup model leads to:

$$\Delta h = \frac{1.3F}{u \, u_*^2} \, (1 + h/\Delta h)^{2/3}$$

where the friction velocity u_* is obtained from the logarithmic wind speed profile law. The default roughness length used in computing u_* is 0.3 m. In convective conditions the breakup model leads to:

$$\Delta h = 4.3 \, (F/u)^{3/5} \, H_*^{-2/5}$$

where H_* is the production rate of turbulence due to convection:

$$H_* = \frac{gQ_0}{C_p T}$$

and Q_0, the sensible heat flux, is set equal to 40 percent of the insolation rate at the surface.

For convective conditions the touchdown model leads to:

$$\Delta h = \frac{F}{uw_d^2} (1 + 2h/\Delta h)^2$$

where $w_d = 0.4w_* = 0.4 (H_* L)^{1/3}$. Note that L is the depth of the mixing height.

Stack-tip downwash is available as an option, as in CRSTER.

The dispersion coefficients used in PPSP are those formulated by Briggs (Gifford, 1975). This formulation is similar to that used in CRSTER in that distinct curves of $\sigma_z(x)$ and $\sigma_y(x)$ are associated with each stability class. For nighttime PPSP uses the same stability class definitions as CRSTER, but during the day the classes A-D are selected on the basis of a convective scaling parameter, u/w_*. As above, w_* is related to the insolation rate. Note that the wind speed used in PPSP for stability classes A through C is the value at one-tenth the depth of the mixing layer. The ranges of u/w_* associated with stability classes A through D are respectively, < 3.5, $3.5 - <6.0$, $6.0 - <13.9$, and ≥ 13.9. Buoyancy-induced dispersion is included in the PPSP as in MPTER.

HPDM. The final rise of a buoyant plume during stable-windy meteorological conditions is the same as that used in CRSTER, but the expression for the stable-calm meteorology differs in that the multiplicative constant is 4.0 rather than 5.0. In either condition, however, the resulting stable-calm plume rise is compared with the neutral/unstable plume rise from Briggs's (1970) equation used in CRSTER and with the neutral-breakup plume rise (Briggs, 1975). The minimum height of final rise that results from these three plume rise estimations is used as the predicted plume rise. During neutral/unstable conditions, the final plume rise used as the predicted rise is the lesser of that obtained from Briggs's (1970) equation used in CRSTER and Briggs's (1975) breakup and touchdown equations used in PPSP.

In each of these calcuations the wind speed and the potential temperature gradient (where applicable) are obtained at an elevation halfway between the top of the stack and the height of final rise. Because the plume rise is initially unknown, an iterative method is used.

Dispersion coefficients are computed from turbulence data. For nonconvective conditions when the Monin-Obukhov length scale lies outside the range -100 m to 0 m, σ_y and σ_z are given by:

$$\sigma_y = i_y \times (1. + 0.9 (t \cdot 15{,}000 \text{ s})^{\frac{1}{2}})^{-1}$$

$$\sigma_z = i_z \times (1 + t/2T_{Lz})^{-1/2}$$

where x is the downwind distance from the source, t is the time of travel, and T_{Lz} is the Lagrangian time scale. Formulas for T_{Lz} are functions of z, L, σ_w, the buoyancy parameter S, and the mixing height. When measured values are not available, i_y and i_z are estimated as functions of u_*, w_*, z, and the Monin-Obukhov length.

Buoyancy enhancement of σ_y and σ_z is included as the sum of variances, and it includes an adjustment for partial penetration of the mixing lid:

$$\sigma b^2 = (\Delta h^2/12.25)(1 - P)^2$$

where P is the penetration factor.

For convectively unstable conditions ($-100 < L < 0$), a probability distribution function (pdf) method is used in place of the Gaussian plume method for the vertical distribution of material. The pdf model simulates the ensemble-averaged concentration field by a superpositioning of the vertical displacements due to source buoyancy and those random vertical velocities due to rising buoyant air and compensating subsidence. The ground-level concentrations are calculated as the sum of contributions from the "direct" plume (plume segments that are carried directly to the surface by convective downdrafts), and from the "indirect" plume (plume material that just impinges on the mixing height and then diffuses to the surface). Plume segments with large values of the buoyancy flux that impinge on the mixing lid remain near the top of the layer for some time and resist downward dispersion over relatively long distances.

TERRAIN FORMULATIONS

Specific adjustments can be made in CRSTER (CEQM), MPTER, SHORTZ, and RTDM to alter the concentrations at receptors located in elevated terrain. TEM, PPSP, and HPDM contain no such adjustment, and are applicable to flat-terrain sites only.

CRSTER (CEQM). CRSTER is considered an "uneven terrain" model in that an adjustment is made to plume height for receptors located in elevated terrain. An adjustment is also made for plume height in situations where stack-tip downwash depresses the plume centerline below the top of the stack. These adjustments entail raising the ground plane (the lower reflecting surface) up to the elevation of the receptor, for each receptor in turn, and raising the mixing height by the same amount. The effect is equivalent to keeping the mixing height fixed, keeping the receptors at the elevation of the plant, and lowering the plume height. Note that receptors that lie below the elevation of the plant are treated as if they were at the plant elevation. CRSTER does not handle receptors located above the elevation of the top of the stack.

MPTER. The influence of terrain is modeled in MPTER by lowering the plume centerline height, as in CRSTER, but not necessarily by the full difference in elevation between the receptor elevation and stack base elevation. Terrain adjustment factors between zero and one may be specified for each stability class. A factor of zero produces an adjustment that is equivalent to CRSTER terrain treatment. A factor of one makes the plume centerline follow the terrain, so that predicted concentrations are the same that would occur in flat terrain. The default factor is zero for all stabilities.

SHORTZ. In SHORTZ, when complex terrain is present, adjustments affect plume heights lying between the receptor and the mixing height. Concentrations are computed only for those plumes that lie in the mixed layer and at those receptors that lie at elevations within the mixed layer. In determining which plumes and receptors are so treated, both the plume height and the mixing height are held at constant elevations above sea level (both heights are considered level surfaces). Once the model decides to make a concentration estimate, the plume height and mixing height are adjusted as in CRSTER, but receptors above the plume centerline height are treated as though they are at the height of the plume. There is no restriction on the height of terrain relative to stack height.

RTDM. For receptors on elevated terrain, RTDM adjusts not only the effective plume height above the surface but also the depth of the mixed layer (stability classes A-D), plus it introduces an overall concentration reduction factor that is called a reflection factor. RTDM also computes a critical dividing-streamline height, H_{crit}, for stability classes E and F, and plumes below H_{crit} are allowed to impinge on terrain.

During neutral and unstable conditions, H_{crit} is zero, and the mixing height is nonzero. In this case both the plume height and the mixing height are adjusted as follows:

$$z_a = C - z \ , \qquad z < z_r$$

$$\quad = z - (1 - C)\, z_r \ , \qquad z > z_r$$

where z denotes either the plume height or the mixing height, z_a denotes the adjusted value, and z_r is the elevation of the terrain at the receptor. C is the plume path coefficient, and is set equal to 0.5 for all stabilities in the default mode of operation. Similar alterations are made to the plume height and the mixing height at several locations upwind of the receptor, and the crosswind-integrated concentration is computed at several heights to obtain the peak concentration at each of these distances. The overall minimum of these peak crosswind-integrated concentrations

(CWIC) is identified and a reflection factor that reproduces the minimum CWIC at the receptor position is computed. The reflection factor that is used is set to the initial reflection factor estimate or 2.0, whichever is less.

During stable conditions, a distinction is made between plumes that are above or below H_{crit}. Above H_{crit}, the adjustment described above is modified by subtracting H_{crit} from both z and z_r. Below H_{crit}, the plume path coefficient is set equal to zero, so that the plume is placed on the surface if the plume height is less than the terrain height, or a CRSTER-like treatment is employed if the plume height is greater than the terrain height.

TREATMENT OF CALM CONDITIONS

All of the models are plume models so that all are formulated for continuous releases of pollutants and an assumed nonzero steady transport wind over the period of the simulation. Regardless of the transport wind speed, receptors at all distances downwind of the source are affected by the plume. Under calm or very light wind speeds, these basic assumptions necessary to the plume model formulation are inappropriate.

If the winds are truly "calm," i.e., the wind speed is below the threshold speed of the sensor, most of the models will use the wind direction from the previous hour, and a minimum wind speed of 1 m/s. The PPSP model simply excludes calm hours from daily and annual averages. To comply with the EPA-suggested minimum data recovery rate of 75 percent, the number of non-calm hours in any averaging period must exceed 75 percent, or else the concentration for the period is set to zero.

If the winds are light and variable, but exceed the threshold of the sensor, then the models that make use of observed turbulence data are better able to simulate this class of meteorology. Nonetheless, even these models are incapable of simulating the non-Gaussian distributions of wind fluctuations, and are unlikely to perform well for variable winds. The model that offers the best formulation for very light wind speeds under convective conditions is the HPDM model with its pdf formulation for convective conditions. Its non-Gaussian formulation for plume behavior in the vertical goes furthest toward representing the physical behavior of the light-wind, convective boundary layer.

4 OVERALL PERFORMANCE OF THE MODELS

The major purpose of evaluating the overall performance of a model is to establish a baseline of model performance for observed and predicted relative concentrations (χ/Q) that are matched by time. In this study we are primarily interested in why certain models perform as they do at the 1-hour averaging period. We have focused our attention on worst-case meteorological conditions, i.e., those that result in the highest concentrations, and how well the model input variables describe such conditions. Thus we have constructed data sets in which the observations and predictions are paired by hour, which will allow us to diagnose why a particular over- or underprediction occurred. In this section we will (1) examine the level of model performance for the highest concentrations and (2) describe the meteorological conditions that seem to be associated most frequently with the largest observations, as well as the largest overpredictions and underpredictions.

THE DATA SETS

In Section 2 we selected four data sets for further model evaluation studies using the following models

Kincaid Study - CEQM, TEM, HPDM
Bull Run Study - CEQM, SHORTZ, RTDM, HPDM
Clifty Creek Study - MPTER/CRSTER
Maryland PPSP Study - CRSTER, PPSP

We developed data sets containing model inputs, model concentrations, and observed concentrations for each of the evaluation studies and for each model. These data sets were all prepared with the same units and variables that could be compared across studies. Because of some sample size limitations not all data sets could be prepared, as we shall discuss later.

In order to diagnose why large residuals occur, we made an effort to develop data sets for all model evaluation exercises with the following variables for the hours when the largest observed and predicted concentrations occur.

Stack data--emission rate, buoyancy flux, and stack height

Model input data--10 m wind speed, mixing height, and PGT stability class

Meteorological and turbulence data--100 m wind speed, friction and convective scaling velocities, and Monin-Obukhov length

From these data we were able to create other useful stability estimators and scaling parameters, such as the scaled buoyancy flux.

Our model evaluation was based on an examination of concentrations normalized to a unit emission rate. This was done to allow for easy intercomparison between sources of different sizes and to remove much of the variation in the source due to changes in plant load. We note, however, that the variations due to plant load have not been removed entirely since changes in load are correlated with buoyancy flux, and hence plume rise and ground-level concentrations. Because of this effect, low load conditions are generally translated into higher normalized ground-level concentrations.

In order to study the meteorological conditions associated with the highest observed concentrations, we made up the data sets for this study using only the 50 hours with the highest observed concentrations and the time-matched largest predictions produced by the models. Since we are interested in those hours when the largest magnitude residuals occur, we also selected the 50 hours with the largest over- or underpredictions. It was reasoned that choosing 50 hours would result in data sets large enough to perform a number of stratifications by meteorology, yet represent a sufficiently high percentile of the cumulative distribution function (CDF).

EPRI PMV&D Data

From the EPRI PMV&D model applications to the Kincaid plant we constructed the following data sets:

On-site data similar to NWS data used as model input

    CEQM predictions and observations of hourly SO_2 data
    CEQM predictions and observations of hourly SF_6 data
    TEM predictions and observations of hourly SO_2 data
    TEM predictions and observations of hourly SF_6 data

Best estimate of model inputs based on on-site data

    CEQM predictions and observations of hourly SO_2 data
    CEQM predictions and observations of hourly SF_6 data

TEM predictions and observations of hourly SO_2 data
TEM predictions and observations of hourly SF_6 data

A data set for HPDM was available but it is different from all other data sets in the sense that (1) only tables of the 25 highest observed and predicted concentrations were available, and (2) the predictions and observations are unpaired in time. It is therefore difficult to directly infer why HPDM under- or overpredicted. To deal with this problem, we made up two data sets for HPDM. One includes the highest predictions and observations unpaired in time; a second, limited set includes only those hours when the highest predictions and observations were matched in time. At the time of this report the complete model evaluation results were not available. These two data sets were prepared for both the SO_2 and the SF_6 observations but had a sample size much less than 50 hours. Therefore we make few if any inferences from the HPDM data sets because of the lack of pairing and small size.

Indirectly observed variables, such as the friction and convective scaling velocities, were estimated using the observed wind profile and an estimate of the sensible heat flux based on observations of net radiation. The sensible heat flux, and consequently the Monin-Obukhov length, was estimated using a variation of the heat budget method described by Holtslag and van Ulden (1983). The methods by which these variables were estimated have been described in detail by Hanna, Weil, and Paine (1986). For Kincaid applications that used on-site meteorological data similar to NWS data, the mixing height was determined from on-site twice-daily T-sonde measurements using the Holzworth method as implemented in the CRSTER/RAMMET meteorological preprocessor. For Kincaid applications that used best estimates of model inputs, the mixing height was either taken directly from T-sonde or acoustic sounder data or predicted using a diagnostic formula like that described by Hanna, Weil, and Paine (1986). When all else failed, the default estimate was obtained via the method used for NWS equivalent data inputs.

From the PMV&D applications to the Bull Run plant we constructed the following data sets of best estimates of model inputs based on on-site data:

CEQM predictions and observations of hourly SF_6 data
RTDM predictions and observations of hourly SF_6 data
SHORTZ predictions and observations of hourly SF_6 data

Two data sets similar to those above were prepared for HPDM, but they have a much smaller sample size (<25).

Maryland PPSP Data

From the Maryland PPSP model evaluation study we were able to construct two data sets: one for CRSTER/MPTER and one for the PPSP model developed by Weil and

Brower (1982). The methods used to estimate the scaling velocities and the Monin-Obukhov length are similar to those used for the PMV&D data sets.

EPA Rural Model Evaluation Data

A review of the 1975 and 1976 Clifty Creek data sets indicated that the meteorology of the two years was similar so we pooled the two years into a single data set. The Clifty Creek data set is different from the data from the other studies in that only the basic meteorological variables used for input data were available; no estimates of the scaling velocities and Monin-Obukhov length were available.

THE HIGHEST OBSERVATIONS

Earlier model performance evaluation studies have shown that observations and predictions matched by time are not well correlated. Our study confirms this observation. The scatter plot in Figure 4-1 illustrates typical model performance under what can be considered, on the basis of observed concentrations, as the worst-case dispersion conditions. As noted by Weil and Brower (1984), models such as CRSTER/MPTER often fail entirely to predict any nonzero ground-level concentration, representing the largest underpredictions by the model.

Some investigators have blamed the poor correlation described above on uncertainties in specifying the proper wind direction, combined with the limitations of a sparse observation network. While this may be true when modeling such places as Clifty Creek, it does not appear to hold for a site like Kincaid. At Kincaid an extensive ground-level tracer network provided coverage sufficiently dense to capture the plume whenever it impacted the ground (conditions where complete plume footprint capture did not occur were generally screened out).

There was no attempt during the PMV+D model evaluation exercise to shift the predicted plume centerline so that predicted concentrations agreed more closely with the observed maximum concentration during each hour. The same model applied to two sites, Kincaid and Clifty Creek (Figure 4-2), produced similar degrees of scatter. This suggests that while uncertainty in wind direction may play a role in producing such poor correlations, it does not appear to be the dominant factor. A quantitative test of this hypothesis would require shifting the receptor array for predictions around to see to what degree predictions of hourly maximums could be improved.

In several evaluation studies all the models under review were shown to have similar performance statistics (Table 4-1). The only major exception is PPSP, which shows a slight overprediction. This is partially due to the fact that the observation network was designed to track the hourly maximum concentration by the use of remote sensing data. The PPSP results summarized in Table 4-1 are very encouraging; however,

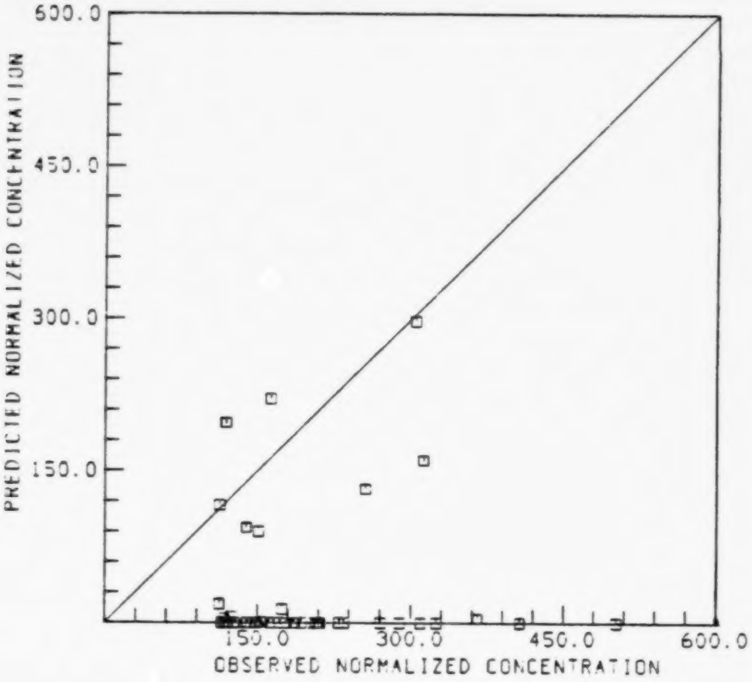


FIGURE 4-1. A scatter plot of predictions and observations for the 50 hours when the highest concentrations were observed at Kincaid. The predictions are for relative SF_6 concentrations made by CEQM. Units are 10^{-9} s/m^3.

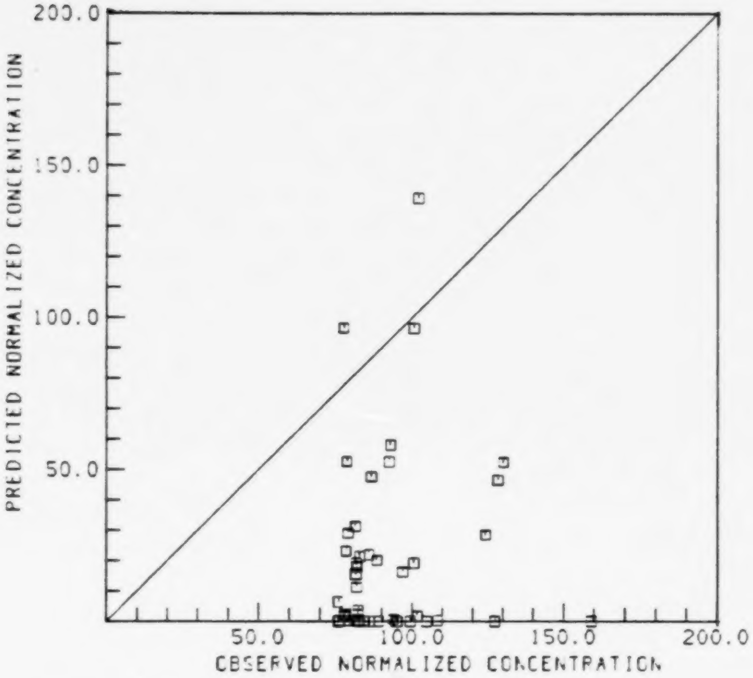


FIGURE 4-2. A scatter plot of predictions and observations for the 50 hours when the highest concentrations were observed at Clifty Creek. The predictions are for normalized SO_2 made by CRSTER/MPTER. Units are 10^{-9} s/m^3.

other model performance evaluations, such as the EPA rural evaluation study, showed that the model produced excessive overpredictions at other sites.

For most of the models under review the average absolute error (AAE) is the same size as the average residual, which in turn is generally in excess of half the magnitude of the average observed concentration. This is consistent with the scatter plots, which show that for about half the hours when a large concentration was observed, a zero concentration was predicted. The degree of underprediction is best illustrated by two other measures listed in Table 4-1, the percent within a factor of two (PWF2) and the fractional difference 2(O - P)/(O + P). The average fractional difference typically approaches one. The fraction of predictions within a factor of two of the observations is typically less than a third (0.33). The only other model besides the PPSP that does slightly better is the RTDM, which posesses some advanced features (see Section 2).

LARGEST OVERPREDICTIONS

Statistics on the largest overpredictions from the 50 hours with the largest magnitude residuals are presented in Table 4-2. It should be noted that of all models and all model performance evaluation studies reviewed by us, only the Clifty Creek CRSTER and the Maryland PPSP model evaluations produced more overpredictions than underpredictions. Scatter plots of the largest overpredictions and underpredictions, like the one in Figure 4-3, reveal that the largest overpredictions often occur when the observed ground-level concentrations are near background. All of the models, with the exception of PPSP, are generally unable to produce concentration predictions within a factor of two of the observations on an hour-to-hour basis. The PPSP model is the only model that avoids several of the biases inherent in the other model predictions (Figure 4-4). One can conclude that either the models are using inaccurate or incomplete input data, or else the dispersion conditions which the models were designed to describe do not happen very often.

In the Kincaid data none of the overpredictions by any model are within a factor of two of the corresponding observations; overpredictions are generally much larger than the average observed concentration. The fractional difference, 2(0 - P/)(0 + P), is generally less than -1.0. The only measure that shows some significance is the correlation between the predicted and observed concentrations. For some models and studies this correlation seems to reflect a real tendency for the predictions to increase with increasing observations, as is the case for the normalized SF_6 concentrations predicted by TEM at Kincaid (Figure 4-5).

LARGEST UNDERPREDICTIONS

Statistics on the largest underpredictions from the 50 hours with the largest magnitude residuals are presented in Table 4-3. For most of the models there are

TABLE 4-1. Summary of statistics for time-matched observations and predictions for the 50 hours with the highest observed concentrations. (Units are 10^{-9} s/m^3).

Study/Model	$\overline{\Delta X}$	$\overline{\Delta X}/\overline{X}$	PWF2	Correlation	AAE	$\overline{X}_{obs}$
Clifty Creek/CRSTER	73 (65, 81)	1.32	14	0.12 (-0.17, 0.38)	75	92
Kincaid SO_2/CRSTER	131 (119, 144)	1.57	6	-0.08 (-0.36, 0.20)	133	150
Kincaid SF_6/CRSTER	165 (140, 190)	1.50	16	0.07 (-0.21, 0.34)	171	192
Bull Run SF_6/CRSTER	113 (102, 161)	0.76	32	0.15 (-0.13, 0.41)	148	214
Maryland PPSP/CRSTER	112 (74, 152)	0.89	30	0.04 (-0.25, 0.31)	160	182
Kincaid SO_2/TEM	128 (117, 139)	1.49	6	-0.07 (-0.35, 0.21)	128	150
Kincaid SF_6/TEM	150 (122, 176)	1.27	12	0.00 (-0.28, 0.28)	163	192
Bull Run SF_6/RTDM	116 (81, 151)	0.74	46	0.30 (0.02, 0.53)	141	214
Bull Run SF_6/SHORTZ	163 (139, 187)	1.22	12	-0.07 (-0.34, 0.21)	172	214
Maryland PPSP/PPSP	-40 (-82, 1)	-0.20	78	0.47 (0.22, 0.66)	96	182

Key

$\overline{\Delta X} = \overline{X}_{obs} - \overline{X}_{pred}$

$\overline{X} = (\overline{X}_{obs} + \overline{X}_{pred})/2.0$

PWF2 = percent within a factor of 2

(,) = 95% confidence intervals

AAE = absolute average error

TABLE 4-2. Summary of statistics for time-matched observations and predictions for hours with the largest overpredictions. (Units are 10^{-9} s/m^3).

Study/Model	No.* Hours	$\overline{\Delta X}$	$\overline{\Delta X}/\overline{X}$	PWF2	Correlation	$\overline{X}_{obs}$
Clifty Creek/CRSTER	42	-111(-116, -105)	-1.62	0	0.52 (0.26, 0.71)	13
Kincaid SO_2/CRSTER	8	-134 (-115, 154)	-1.52	0	0.14 (-0.63, 0.77)	21
Kincaid SF_6/CRSTER	21	-269 (-228, -311)	-1.48	0	-0.04 (-0.46, 0.40)	47
Bull Run SF_6/CRSTER	9	-198 (-134, 262)	-0.97	23	0.56 (-0.17, 0.89)	105
Maryland PPSP/CRSTER	13	-178 (-102, -253)	-1.04	8	0.85 (0.56, 0.95)	82
Kincaid SO_2/TEM	8	-126 (-108, -143)	-1.84	0	0.82 (0.27, 0.97	5
Kincaid SF_6/TEM	7	-146 (-89, -203)	-1.05	0	0.87 (0.33, 0.98)	66
Bull Run SF_6/RTDM	4	-201 (-4, -397)	-0.73	25	0.82 (-0.66, 0.99)	174
Bull Run SF_6/SHORTZ	5	-178 (-81, -275)	-1.04	20	0.84 (-0.17, 0.99)	81
Maryland PPSP/PPSP	41	-139 (-95, -184)	-0.82	39	0.66 (0.44, 0.81)	99

* nonzero predictions and observations only

Key

$\overline{\Delta X} = \overline{X}_{obs} - \overline{X}_{pred}$

$\overline{X} = (\overline{X}_{obs} + \overline{X}_{pred})/2.0$

PWF2 = percent within a factor of 2

(,) = 95% confidence intervals

AAE = absolute average error

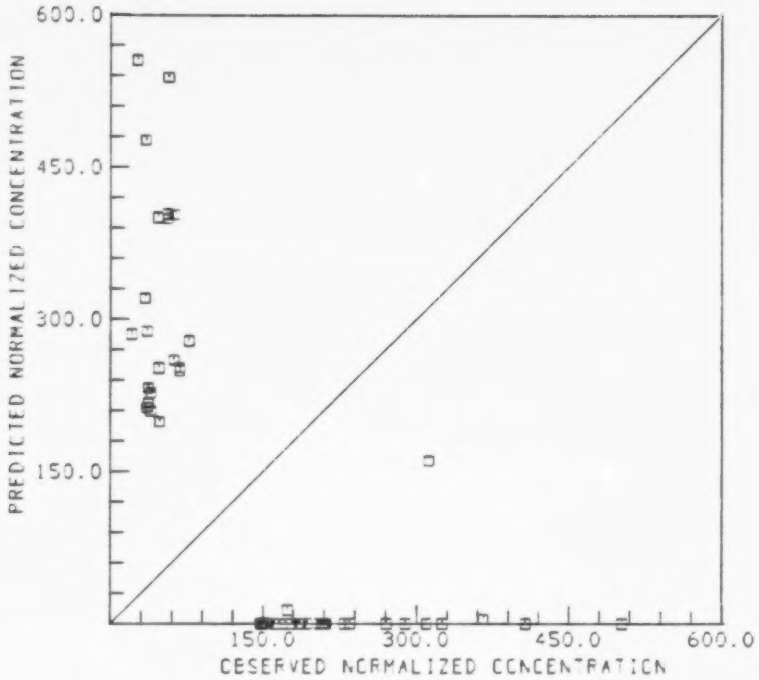


FIGURE 4-3. A scatter plot of the 50 largest overpredictions and underpredictions of normalized SF_6 made by CEQM at Kincaid. The units are 10^{-9} s/m^3.

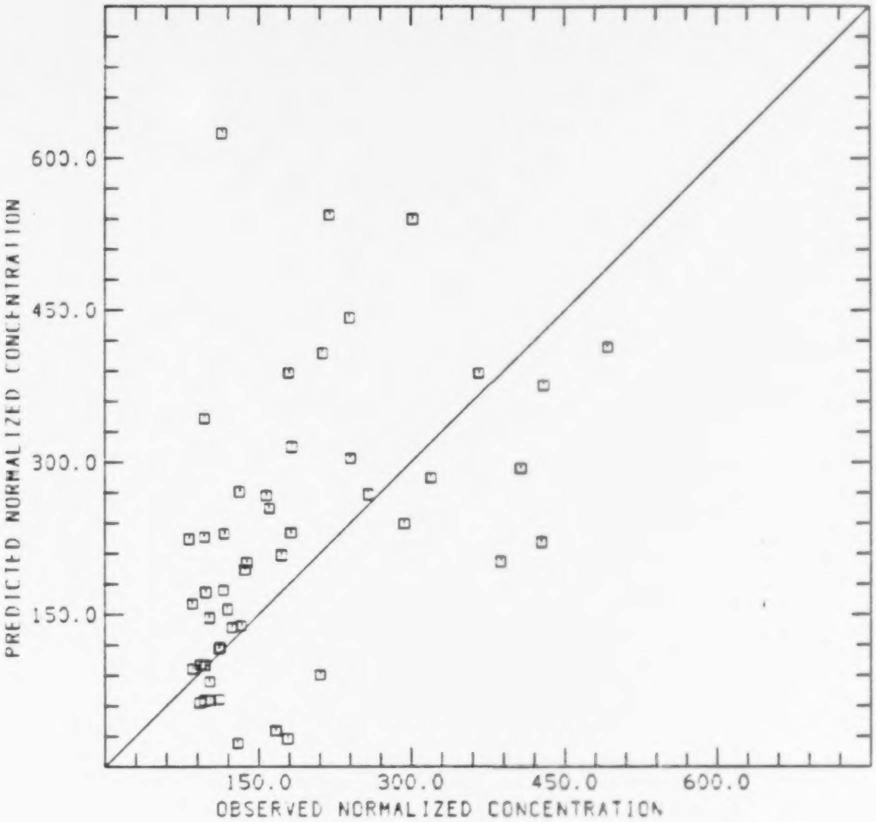


FIGURE 4-4. A scatter plot of observations and predictions for the 50 hours with the highest observed concentrations. The predictions are for normalized SF_6 made by the PPSP model for the PPSP data set. Units are 10^{-9} s/m^3.

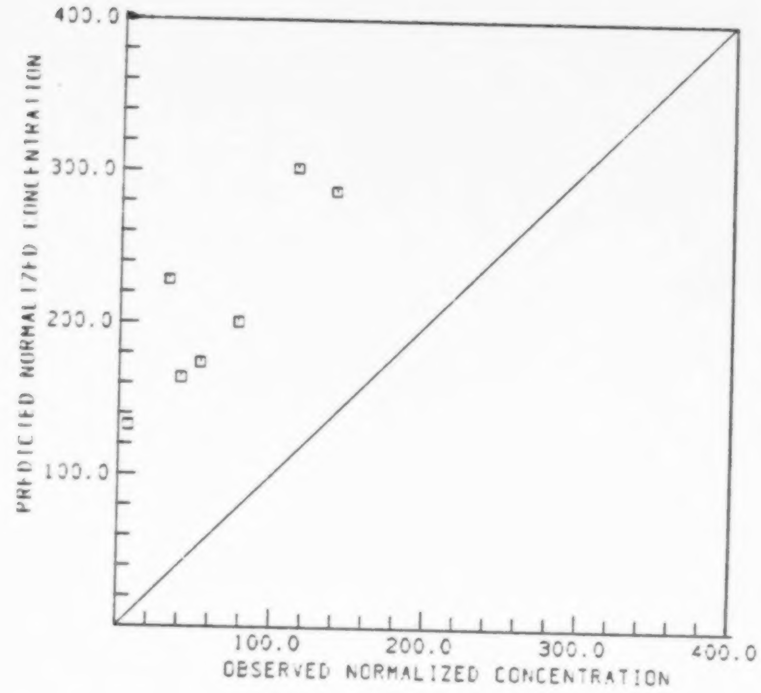


FIGURE 4-5. A scatter plot of the largest 7 overpredictions by TEM of normalized SF_6 concentrations at Kincaid. The units are 10^{-9} s/m^3.

TABLE 4-3. Summary of statistics for time-matched observations and predictions for hours with the largest underpredictions. (Units are 10^{-9} s/m^3).

Study/Model	No.* Hours	$\overline{\Delta X}$	$\overline{\Delta X}/\overline{X}$	PWF2	Correlation	$\overline{X}_{obs}$
Clifty Creek/CRSTER	7	113 (96, 130)	1.85	0	0.11 (-0.70, 0.80)	117
Kincaid SO_2/CRSTER	42	147 (135, 159)	1.87	0	0.22 (-0.10, 0.48)	152
Kincaid SF_6/CRSTER	29	223 (194, 252)	1.90	3	0.18 (-0.20, 0.51)	228
Bull Run SF_6/CRSTER	40	180 (154, 206)	1.58	0	0.50 (0.22, 0.70)	204
Maryland PPSP/CRSTER	19	184 (122, 247)	1.42	15	0.43 (-0.03, 0.74)	222
Kincaid SO_2/TEM	42	140 (128, 152)	1.73	0	0.19 (-0.12, 0.46)	151
Kincaid SF_6/TEM	43	181 (158, 204)	1.75	0	0.22 (-0.08, 0.49)	194
Bull Run SF_6/RTDM	45	167 (142, 193)	1.72	5	0.55 (0.31, 0.73)	181
Bull Run SF_6/SHORTZ	44	188 (166, 211)	1.59	0	0.30 (0.0, 0.55)	213
Maryland PPSP/PPSP	9	127 (0.0, 255)	0.60	45	0.95 (0.78, 0.99)	274

* nonzero predictions and observations

Key

$\overline{\Delta X} = \overline{X}_{obs} - \overline{X}_{pred}$
$\overline{X} = (\overline{X}_{obs} + \overline{X}_{pred})/2.0$
PWF2 = percent within a factor of 2
(,) = 95% confidence intervals
AAE = absolute average error

generally more underpredictions than overpredictions. Furthermore, during these hours of underpredictions one does not find as much of a systematic association between observations and predictions as is observed for the hours of overpredictions. Part of this loss of association seems to be due to the fact that the models (with the exception of PPSP) predict zero concentration for a wide range of observed concentrations (Figure 4-6).

We can note from Table 4-3 that for most models the average underprediction, Δx, is slightly smaller than the average observed concentration. With the exception of the Maryland PPSP study data set, almost all of the predictions are smaller than the observations by more than a factor of two. The fractional difference, $2(0 - P)/(0 + P)$, is typically greater than 1.5 and in some cases approaches the maximum possible value of 2.0.

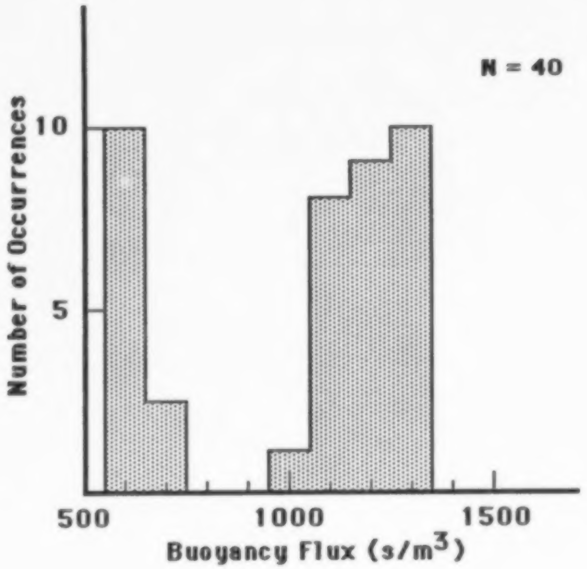
METEOROLOGY ASSOCIATED WITH HIGHEST CONCENTRATIONS

One can make the following observations about the meteorology associated with the 50 hours with the highest observed concentrations (see Table 4-4):

> Several types of meteorological conditions occur. The Bull Run site with its low wind speeds, high mixing heights, and highly unstable conditions appears to have the most unusual meteorology.
>
> At Bull Run the meteorology most often associated with the highest concentrations is very low wind speed, high mixing heights, and very unstable conditions.
>
> At all other sites the most frequent meteorology associated with the highest concentrations includes 2-4 m/s wind speeds, mixing heights between 500 and 1000 m, and either unstable or neutral conditions.
>
> For all studies reviewed the convective scaling velocity is approximately 2 m/s.

The average buoyancy flux ranges by a factor of 10 among the sites under review; however, for the 50 hours with the largest observed concentrations at each site it is usually substantially less than the average. Together, these observations suggest that unstable plume trapping is the most frequent dispersion condition associated with the highest observed concentrations.

The largest underpredictions tend to occur most often under unstable conditions with low to moderate wind speeds (Table 4-5). At the Clifty Creek site they tend to occur



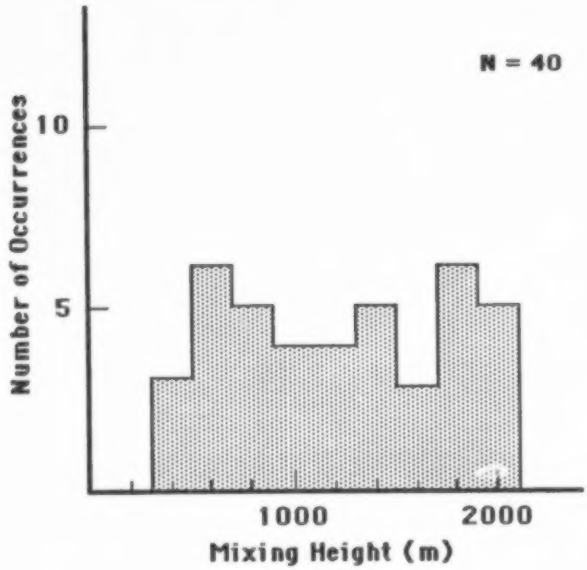


FIGURE 4-6. Values of buoyancy flux and mixing height associated with underpredictions made by CEQM at the Bull Run site.

TABLE 4-4. Ranges for meteorological variables that occur most frequently during the 50 hours with the highest concentrations. The numbers in parentheses denote the sample size; all units are MKS.

Study/Model	W_s (10 m)	Mixing Height	PGT Stability Class	Buoyancy Flux	w_*
Clifty Creek/CRSTER	2.5-3.5 (12)	600-1000 (12)	4 (24)	125-135 (10)	--
Kincaid SO_2/CRSTER	3-5 (19)	600-1000 (15)	2 (21)	900-1100 (18)	2.2-2.6 (15)
Kincaid SF_6/CRSTER	2.5-3.5 (12)	700-900 (14)	2 (23)	900-1100 (19)	1.75-2.25 (13)
Bull Run SF_6/CRSTER	1.25-1.75 (18)	1700-1900 (9)	2 (35)	550-650 (12)	2.1-2.3 (11)
Maryland PPSP/CRSTER	2.5-3.5 (13)	700-900 (10)	4 (22)	300-500 (28)	1.9-2.1 (14)
Kincaid SO_2/TEM	3-5 (19)	600-1000 (15)	2 (21)	900-1100 (18)	1.75-2.25 (17)
Kincaid SF_6/TEM	2.5-3.5 (12)	600-1000 (14)	2 (23)	900-1100 (19)	1.75-2.25 (13)
Bull Run SF_6/RTDM	1.25-1.75 (18)	1700-1900 (9)	1 (32)	550-600 (12)	2.1-2.3 (11)
Bull Run SF_6/SHORTZ	1.25-1.75 (18)	1700-1900 (9)	1 (32)	550-650 (12)	2.1-2.3 (11)
Maryland PPSP/PPSP	2.5-3.5 (13)	700-900 (10)	4 (22)	300-500 (28)	1.9-2.1 (14)

TABLE 4-5. Ranges for meteorological variables that occur most frequently during hours with the largest underpredictions. The numbers in parentheses denote the sample size; all units are MKS.

Study/Model	W_s (10 m)	Mixing Height	PGT Stability Class	Buoyancy Flux	w_*
Clifty Creek/CRSTER	0.5-1.5 (3)	500-700 (2)	3 (4)	185-195 (2)	--
Kincaid SO_2/CRSTER	2.5-3.5 (12)	500-700 (13)	2 (14)	900-1100 (14)	2.2-2.6 (12)
Kincaid SF_6/CRSTER	3.25-3.75 (8)	0-200 (8)	2 (10)	900-1100 (12)	1.75-2.25 (10)
Bull Run SF_6/CRSTER	1.3-1.5 (12)	500-700 (6)	2 (32)	550-650 (10)	2.1-2.3 (8)
Maryland PPSP/CRSTER	1.0-3.0 (13)	700-900 (9)	4 (17)	300-500 (19)	1.9-2.1 (10)
Kincaid SO_2/TEM	2.5-3.5 (12)	500-700 (14)	2 (15)	900-1100 (15)	1.75-2.25 (16)
Kincaid SF_6/TEM	3.25-3.75 (10)	200-600 (15)	2 (18)	900-1100 (19)	1.75-2.25 (10)
Bull Run SF_6/RTDM	1.1-1.3 (11)	1500-1700 (7)	1 (24)	1250-1350 (14)	2.1-2.3 (10)
Bull Run SF_6/SHORTZ	1.1-1.3 (12)	1700-1900 (8)	1 (27)	1250-1350 (12)	2.1-2.3 (10)
Maryland PPSP/PPSP	2.75-3.25 (2)	700-900 (3)	4 (5)	275-300 (2)	1.9-2.1 (3)

under convectively unstable, near-calm conditions with low mixing heights and higher than normal buoyancy fluxes. A similar dispersion regime occurs at Bull Run for RTDM and SHORTZ, with the exception that the buoyancy fluxes and mixing heights are higher than even the average of these variables for the 50 hours with the largest magnitude residuals. A histogram of the buoyancy flux for the hours with the largest underpredictions at the Bull run site is bimodal, while the histogram of mixing height for the same hours is nearly uniform (Figure 4-6). The bimodality in the buoyancy flux explains in part why the CRSTER underpredictions for the Bull Run site occur most often at low buoyancy fluxes, since all it takes is a small shift in the number of cases to move the peak from small buoyancy fluxes to large ones.

For hours with the largest underpredictions at the Kincaid site, moderate wind speeds, low or moderate mixing heights, and unstable conditions occur most frequently. At most sites the buoyancy fluxes are most often average. The Maryland PPSP study indicates that low to moderate wind speed, neutral conditions, and moderate buoyancy fluxes are the most frequent conditions. Despite differences in w_* between sites, the most frequently occurring convective wind speed falls in the range of 1.8 - 2.2 m/s, which is consistent with the w_* for the 50 hours with the highest observed concentrations.

Table 4-6 presents the most frequently occurring ranges of a number of meteorological variables for those hours when the largest overpredictions occur. Note that the sample sizes for some sites are so small it is difficult to pick out the most frequently occurring conditions; sometimes there is no maximum, or even a simple tie for the maximum concentration. This is why blanks occur in Table 4-6. At Kincaid the largest overpredictions occur most frequently under moderate wind speed conditions, unstable conditions, and low to moderate mixing heights. Two-way tables of residuals versus buoyancy flux suggest a variable and inconclusive relationship between overprediction and buoyancy flux. At Bull Run the largest overpredictions occur most frequently during near-calm winds in conjunction with very unstable conditions.

At the Clifty Creek site large overpredictions are produced most often during hours when moderate wind speeds and low buoyancy fluxes occur. The most frequently occurring stability class for the Clifty Creek and Maryland PPSP studies when large overpredictions occur was either slightly unstable or neutral conditions. High wind speeds and low buoyancy fluxes, typical of low plume rise conditions, were most prevalent for the PPSP model overpredictions; low wind speeds were most frequent for CRSTER overpredictions during the same study.

The most notable features of this brief review of meteorology are that (1) several dispersion regimes seem to be responsible for large over- and underpredictions and (2) the frequency of the dispersion conditions varies both between sites and between models at the same site. This suggests that pooling sample sets for a specific model or for several models for purposes of analysis may produce some oversimplifications

TABLE 4-6. Ranges for meteorological variables that occur most frequently during hours with the largest overpredictions. The numbers in parentheses denote the sample size; all units are MKS.

Study/Model	W_s (10 m)	Mixing Height	PGT Stability Class	Buoyancy Flux	w_*
Clifty Creek/CRSTER	2.25-2.75 (13)	900-1100 (10)	3 (23)	145-155 (12)	--
Kincaid SO_2/CRSTER	3.25-3.75 (3)	500-700 (3)	2 (3)	900-1100 (2)	2.3-2.5 (2)
Kincaid SF_6/CRSTER	2.25-2.75 (3)	200-600 (9)	2 (9)	500-700 (6)	2.85-2.95 (6)
Bull Run SF_6/CRSTER	1.1-1.3 (3)	1500-1700 (2)	1 (8)	1250-1350 (5)	1.9-2.1 (3)
Maryland PPSP/CRSTER	1.5-2.5 (12)	500-700 (6)	4 (24)	350-450 (11)	1.9-2.1 (10)
Kincaid SO_2/TEM	2.5-3.5 (3)	500-700 (2)	3 (4)	550-650 (3)	1.7-1.9 (3)
Kincaid SF_6/TEM	2.25-2.75(4)	900-1100 (3)	2 (4)	500-700 (3)	1.5-1.7 (2)
Bull Run SF_6/RTDM	1.05-1.15 (2)	--	1 (3)	570-590 (2)	2.3-2.5 (2)
Bull Run SF_6/SHORTZ	--	--	1 (2)	550-650 (4)	1.3-1.7 (2)
Maryland PPSP/PPSP	7.5-8.5 (4)	900-1100 (3)	3 (10)	325-375 (4)	1.9-2.1 (5)

and obscure the relationship between specific meteorological regimes and large-magnitude residuals. Stratifications of the data by site and model will be required in order to resolve meteorological conditions responsible for certain types of over- or underpredictions.

5 MODEL PERFORMANCE AS A FUNCTION OF INPUT DATA

Maximum concentrations from very tall stacks (> 150 m) in flat or uncomplicated terrain have been found to occur most frequently during light-wind convective conditions (Gryning et al., 1987). Stacks that are shorter (e.g., 50 to 100 m) have been found to produce maximum concentrations under high wind conditions, when plume rise is reduced, so that a coning plume impacts the ground significantly (Hanna, Weil, and Paine, 1987). Several factors are important in determining which regime is dominant. An important factor is whether or not the plume is sufficiently buoyant to rise through the mixed layer. If the buoyancy flux is not sufficient, and the wind speeds are light to moderate, the plume is likely to loop and wander. For this study we examined the meteorological criteria needed for each of these three plume behaviors.

Our analysis, which uses convective scaling theory, was performed on meteorological and tracer observations made at the Kincaid and Bull Run sites. The analysis indicates that the maximum concentrations are likely to occur under free convective conditions at both of these sites. (Details of this analysis are presented in the appendix.) We would also expect the maximum concentrations to occur under convective conditions at the three sites in the Maryland PPSP study. The Clifty Creek site, because of low buoyancy flux, is expected to have maximum concentrations typical of moderate wind speed and daytime conditions, or neutral and high-wind conditions.

To obtain statistically significant information on those meteorological conditions when the highest concentrations occur, we conducted a model-by-model, site-by-site exploration of the relation of SO_2 concentrations to variations in the model input variables. The first step in this effort was to conduct a correlation analysis, making extensive use of scatter plots to enable us to make some visual sense of the association of meteorology with the differences between the observations and predictions. In the following discussion of the findings of this analysis, we separately examine the highest 50 observed concentrations, the largest overpredictions, and the largest underpredictions.

ANALYSIS OF HOURLY SO_2 CONCENTRATION DATA—THE KINCAID DATA

Several specialized analyses of the Kincaid data set were conducted in order to understand how certain variables, such as the wind speed and Monin-Obukhov length, relate to maximum observed concentrations. One of the reasons for conducting such an analysis is to determine under what conditions the maximum concentrations occur, and to confirm that the very highest concentrations occur under convective, low wind speed conditions.

Data on the top 25 one-hour SO_2 concentrations at Kincaid data sets are presented in Tables 5-1 and 5-2. These tables present concurrent meteorological and emissions data and the identity of the monitor for each of the 25 concentrations. None of these 50 hours could be classed as "neutral, high wind." Nearly all are associated with light-wind convective periods. The 13 m/s wind speed at the 10 m level on 15 March 1981 is the only wind speed above 10 m/s, and it occurred during a nocturnal jet situation (as is evident from the low wind speed at the 10 m level). The concentration of 263 ppb that occurred at 0200 on 27 April 1980 was associated with nearly neutral conditions ($L = 100$ m), although the wind speed at the 100 m level of the tower was not particularly high (7.2 m/s).

The readings of 379 and 921 ppb at monitor N at 0800 and 0900 on 15 May 1980 were isolated maxima. Even though one high hour followed another in time, the data are questionable because of the spatial discontinuity and the fact that convection was not present.

The full 5208 hours of maximum hourly SO_2 concentrations with concurrent wind speed (U) and Monin-Obukhov length (L) was used to create a joint cumulative density function (CDF) of concentration and meteorological condition. The wind speed at 10 m was chosen to avoid the influence of nocturnal jets at 100 m. The data were grouped into narrow ranges of U and L and the median and other points on the CDF within each range were determined. These results are given in Table 5-3. The concentration of 921 ppb that appeared in Table 5-2 does not appear here; it was removed from the data base because of a deficiency in the meteorological data.

The results for wind speed show that maximum concentrations in the range from 427 to 487 ppb can occur at light to moderate wind speeds, anywhere within the range from 1 to 6 m/s. Lower concentrations, from 100 to 200 ppb, occur for wind speeds greater than 6 m/s. It is interesting, however, that the median concentration is larger at high wind speeds than at low wind speeds. Apparently the maxima under light-wind convective conditions are larger but occur less frequently than those under high-wind neutral conditions.

The results for Monin-Obukhov length corroborate the conclusion that the 1-hour maximum SO_2 concentrations at Kincaid occurs during convective conditions ($-31.6 < L < 0$). However, two hours of maximum concentrations (203 and 262 ppb) occur during neutral conditions ($|L| > 100$ m). It is evident that concentrations are usually low during light wind, stable conditions ($L < 20$ m). Most of the information at the 0 to 50 percent levels of the cumulative distribution functions is probably within the range of uncertainty in the data ($C < 30$ ppb).

TABLE 5-1. Data for top 25 hourly SO_2 concentrations at Kincaid for developmental data set. (Units are g-m-s)

	OBSERVED																	
						10 m			h						PGT			
Rank	CONC PPB	YR	M	D	H	WD	WS	WS Tower 100m	pred.	u_*	w_*	σ_v	σ_w	L	stab class	F s^4/m^3	Q g/s	Monitor id
1	487.0	80	4	25	13	52	2.1	2.3	1332	.22	1.95	.49	.90	-4.8	1	1987	8243	6
2	483.0	80	4	25	14	53	3.0	3.3	1375	.28	2.05	1.11	.77	-10.4	2	2003	8410	6
3	468.0	80	4	25	12	57	2.0	2.3	1156	.22	1.95	1.35	.84	-3.9	1	2027	8097	6
4	427.0	80	4	23	13	36	5.1	6.0	1212	.50	2.32	.71	.78	-30.3	3	2036		10
5	392.0	80	4	25	11	124	1.7	1.6	952	.23	2.02	.96	.71	-3.4	1	2066	7788	T
6	338.0	80	4	23	14	348	-	4.1	1269	.44	2.8	-	-	-27.9	3	1972		10
7	333.3	80	8	9	12	289	1.8	2.0	1402	.26	2.52	1.02	.68	-3.9	1	1460	6700	G
8	319.0	80	7	13	11	52	2.4	2.9	897	.27	2.05	.28	.63	-5.2	1	561	2376	10
9	311.0	81	6	17	15	*												4
10	310.0	80	5	2	11	245	1.8	2.2	1507	.24	2.22	1.26	.67	-4.8	1	2185	8754	C
11	299.9	80	6	18	11	138	1.8	2.0	1173	.26	2.20	1.18	-	-4.7	1	989	4231	T
12	292.0	80	5	2	14	52	1.4	1.7	2043	.16	1.75	1.01	.52	-3.8	1	2116	8326	6
13	287.0	80	5	2	15	61	2.1	2.3	2087	.25	2.20	1.34	.65	-7.9	2	2156	8553	6
14	273.0	80	5	4	13	95	2.0	1.9	2650	.26	2.86	.79	.75	-4.9	1	2235	8844	6
15	268.0	80	5	2	16	48	1.9	2.2	2119	.20	1.99	1.08	.89	-5.7	1	2128	8562	10
16	265.0	80	5	3	10	24	4.6	5.5	639	.47	1.70	.71	.70	-33.4	3	2323	9153	10
17	263.9	80	6	18	12	254	2.0	2.2	1301	.27	2.30	1.13	-	-5.5	1	1014	4324	5
18	251.0	80	7	22	10	324	3.6	4.8	2045	.41	2.69	.55	.73	-18.7	2	1238	5531	1
19	250.0	80	5	4	12	7	2.2	2.5	1911	.28	2.67	.79	.92	-5.8	1	2162	8606	6
20	240.1	80	6	18	13	295	2.0	2.4	1412	.27	2.37	1.44	-	-5.5	1	980	4151	0
21	233.1	80	5	24	14	223	2.1	2.7	1599	.27	2.25	.64	.68	-6.6	1	914	3577	5
22	232.0	80	5	2	17	39	2.3	2.7	2131	.24	1.43	1.11	.49	-25.3	3	2172	8741	10
23	224.6	81	3	15	1	207	3.8	13.0	160	.29	-	.11	-	64	5	824	3716	13
24	223.0	80	4	23	15	29	3.6	4.4	1318	.37	2.11	1.05	.91	-17.9	2	2006		10
25	219.0	80	8	11	12	313	3.1	3.9	976	.38	2.23	.78	.76	-11.9	2	887	4853	7

* 6/17/81 not in KISO2DEV File

TABLE 5-2. Data for top 25 hourly SO_2 concentrations at Kincaid for evaluation data set. (Units are g-m-s)

	OBSERVED																	
						10 m												
Rank	CONC PPB	YR	M	D	H	WD	WS	WS Tower 100m	h pred.	u_*	w_*	σ_v	σ_w	L	PGT stab class	F S^4/m^3	Q g/s	Monitor id
1	921.1	80	5	15	9	*												N
2	464.0	80	6	21	13	147	1.8	1.9	1882	.26	2.68	1.13	.95	-4.4	1	1083	4512	4
3	378.9	80	5	15	8	*												N
4	342.5	80	8	25	12	199	2.8	3.5	1126	.36	2.35	.66	.73	-9.9	1	1430	6094	J
5	300.0	80	7	17	17	191	2.2	2.3	2543	.28	2.38	.32	.15	-10.3	2	1008	4144	1
6	262.8	80	4	27	2	16	4.6	7.2	220	.36	0	.25	.48	100		2029	7080	C
7	255.0	80	6	21	10	165	3.1	4.0	1433	.37	2.29	.71	.65	-15	2	718	3115	D
8	251.2	80	8	25	11	198	3.7	4.6	1050	.43	2.26	.50	1.12	-18.7	2	1497	6269	J
9	248.0	80	4	15	13	281	2.5	6.3	1086	.29	2.09			-7.2	4	2160	8606	C
10	246.0	80	7	29	11	313	3.1	3.8	1031	.38	2.26	.53	.83	-12.2	2	1623	7391	1
11	244.0	80	7	29	12	319	3.6	4.5	1590	.42	2.66	.54	.77	-15.8	2	1754	8192	1
12	238.8	80	8	26	12	192	3.2	3.9	1367	.29	2.51	.36	.57	-13.1	2	1316	5610	J
13	238.0	80	8	26	15	161	3.1	3.9	1996	.38	2.63	.85	.47	-14.4	2	1210	5374	1
14	237.0	80	6	21	15	105	2.2	2.3	1996	.29	2.49	.65	.55	-7.6	2	1065	4540	4
15	234.0	81	4	6	10	158	2.9	3.6	837	.31	1.63	.44	.59	-14.7	2	1004	4225	A
16	230.6	80	8	26	13	195	2.8	3.6	1841	.36	2.77	.83	.57	-9.9	2	1236	5492	L
17	215.0	80	5	6	9	322	4.6	5.8	427	.46	1.35	.35	.50	-42	3	1050	4181	1
18	215.0	80	8	12	10	312	3.3	4.3	1290	.38	2.15	.47	.75	-17.9	2	947	4777	7
19	212.0	80	7	17	11	35	1.6	1.7	1893	.24	2.80	.33	.55	-3.1	1	1411	5515	B
20	207.0	80	8	27	15	172	3.3	4.1	1825	.39	2.88	.50	.55	-17.5	2	1384	5892	2
21	206.1	80	8	26	11	190	3.3	4.3	1009	.40	2.22	.34	.61	-14.6	2	1242	5375	I
22	206.0	80	7	29	14	343	4.4	4.8	2290	.49	2.94	.38	.77	-26.2	2	1769	8469	9
23	205.0	80	6	3	10	*												9
24	204.8	80	8	27	14	169	2.9	3.6	2802	.36	3.05	.58	.55	-11.7	2	1332	5644	L
25	204.0	81	4	6	12	176	3.9	4.8	1679	.40	2.35	1.26	.88	-21.4	3	995	4184	4

* Day not in file KISO2EVL.

TABLE 5.3. Variation of maximum hourly SO_2 concentrations with U and L at Kincaid.

No.	Range	No. Pts	Cumulative Distribution										
			Min	0.5 %	2.5 %	5.0 %	25.0 %	50.0 %	75.0 %	95.0 %	97.5 %	99.5 %	Max
	MAX. CHI (PPB) DISTRIBUTION FOR EACH CATEGORY OF WIND SPEED (M/S)												
1	0.0 <= WS < 1.0	119	1.00	1.00	2.00	3.00	10.00	22.00	29.00	39.00	77.00	99.00	112.00
2	1.0 <= WS < 2.0	843	1.00	2.00	3.00	3.00	18.00	27.00	31.00	102.00	158.00	292.00	464.00
3	2.0 <= WS < 3.0	1131	1.00	2.00	3.00	4.00	19.00	28.00	34.00	131.00	184.00	273.00	487.00
4	3.0 <= WS < 4.0	884	1.00	1.00	2.00	3.00	10.00	26.00	34.00	101.00	164.00	244.00	483.00
5	4.0 <= WS < 5.0	630	1.00	1.00	2.00	3.00	12.00	25.00	36.00	111.00	138.00	187.00	265.00
6	5.0 <= WS < 6.0	450	1.00	1.00	2.00	3.00	7.00	20.00	32.00	102.00	127.00	186.00	427.00
7	6.0 <= WS < 7.0	271	1.00	1.00	3.00	3.00	7.00	20.00	35.00	76.00	88.00	112.00	163.00
8	7.0 <= WS < 8.0	163	1.00	1.00	3.00	4.00	10.00	21.00	31.00	75.00	94.00	117.00	118.00
9	8.0 <= WS < 9.0	116	1.00	1.00	2.00	4.00	13.00	23.00	33.00	68.00	103.00	125.00	132.00
10	9.0 <= WS < 10.0	61	2.00	2.00	2.00	4.00	10.00	20.00	26.00	68.00	81.00	105.00	134.00
11	10.0 <= WS < 11.0	34	1.00	1.00	1.00	1.00	8.00	23.00	42.00	69.00	82.00	82.00	85.00
12	11.0 <= WS < 12.0	31	1.00	1.00	1.00	1.00	4.00	22.00	47.00	77.00	175.00	175.00	203.00
13	12.0 <= WS < 13.0	22	1.00	1.00	1.00	1.00	8.00	28.00	49.00	72.00	113.00	113.00	145.00
14	13.0 <= WS < 14.0	13	11.00	11.00	11.00	11.00	24.00	48.00	77.00	117.00	117.00	117.00	128.00
15	14.0 <= WS < 15.0	7	17.00	17.00	17.00	17.00	17.00	73.00	82.00	84.00	84.00	84.00	86.00
16	15.0 <= WS < 16.0	6	35.00	35.00	35.00	35.00	35.00	62.00	77.00	82.00	82.00	82.00	102.00
17	16.0 <= WS < 17.0	1	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00	115.00
18	17.0 <= WS < 18.0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	18.0 <= WS < 19.0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	19.0 <= WS	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MAX. CHI (PPB) DISTRIBUTION FOR EACH CATEGORY OF MONIN-OBUKHOV LENGTH (M)												
1	-999.0 <= L < -501.2	198	1.00	1.00	2.00	3.00	9.00	22.00	38.00	84.00	104.00	128.00	203.00
2	-501.2 <= L < -251.2	170	1.00	1.00	2.00	3.00	10.00	23.00	36.00	81.00	117.00	145.00	175.00
3	-251.2 <= L < -125.9	279	1.00	1.00	2.00	3.00	12.00	23.00	37.00	88.00	105.00	135.00	195.00
4	-125.9 <= L < -63.1	376	1.00	1.00	2.00	3.00	15.00	27.00	38.00	104.00	124.00	166.00	193.00
5	-63.1 <= L < -31.6	482	0.00	0.00	3.00	4.00	19.00	28.00	47.00	112.00	145.00	194.00	265.00
6	-31.6 <= L < -15.8	462	1.00	2.00	3.00	4.00	23.00	31.00	59.00	148.00	187.00	251.00	427.00
7	-15.8 <= L < -7.9	280	2.00	2.00	4.00	5.00	27.00	36.00	75.00	204.00	238.00	300.00	483.00
8	-7.9 <= L < -4.0	220	1.00	1.00	5.00	10.00	30.00	54.00	121.00	237.00	273.00	319.00	487.00
9	-4.0 <= L < -2.0	65	1.00	1.00	1.00	5.00	26.00	46.00	88.00	217.00	333.00	392.00	468.00
10	-2.0 <= L < 0.0	16	3.00	3.00	3.00	3.00	21.00	24.00	32.00	64.00	64.00	64.00	99.00
11	0.0 <= L < 2.0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	2.0 <= L < 4.0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	4.0 <= L < 7.9	8	3.00	3.00	3.00	3.00	27.00	28.00	33.00	35.00	35.00	35.00	88.00
14	7.9 <= L < 15.8	1109	1.00	2.00	3.00	3.00	10.00	23.00	29.00	36.00	41.00	62.00	96.00
15	15.8 <= L < 31.6	199	0.00	0.00	2.00	3.00	10.00	22.00	30.00	46.00	69.00	91.00	108.00
16	31.6 <= L < 63.1	356	0.00	0.00	2.00	3.00	7.00	21.00	29.00	38.00	64.00	108.00	150.00
17	63.1 <= L < 125.9	348	2.00	2.00	2.00	3.00	7.00	20.00	29.00	40.00	56.00	96.00	262.00
18	125.9 <= L < 251.2	272	1.00	1.00	2.00	2.00	5.00	18.00	23.00	37.00	63.00	73.00	85.00
19	251.2 <= L < 501.2	110	1.00	1.00	3.00	3.00	7.00	15.00	23.00	46.00	67.00	73.00	106.00
20	501.2 <= L	89	1.00	1.00	1.00	2.00	8.00	20.00	29.00	75.00	82.00	87.00	115.00

Total no. data points = 5208 (including missing value indicators)

Wind speed, Max. Chi -- no. valid pts. = 4782
Wind speed, Max. Chi -- no. invalid pts. = 426
Monin-Obukhov length, Max. Chi -- no. valid pts. = 5039
Monin-Obukhov length Max. Chi -- no. invalid pts. = 169

SCATTER PLOTS AND CORRELATION ANALYSIS

Several model input parameters were correlated with the residuals for observations minus predictions for the 50 hours with the highest concentrations. We have already noted that these 50 hours contain more underpredictions than overpredictions. The input parameters that were selected to be correlated against the residuals were:

10 meter wind speed (U(10 m))
Mixing length (h)
Buoyancy flux (F)
PGT stability class (I)

In addition, on the basis of the findings presented in the appendix, we also computed and used the stability parameter, U(10 m)/W_*, plus the scaled buoyancy flux, defined as

$$F_* = F/[U\ (10\ m)\ h\ W_*^2] \quad . \tag{1}$$

where W_* is the convective scaling velocity.

The results of the correlation study are summarized in Table 5-4. None of the models have residuals that significantly correlate with 10 m wind speed. A few model applications, such as SHORTZ at Bull Run, show some correlation of residuals with mixing height. Interestingly enough, this correlation at Bull Run is positive, i.e., the larger the residual, the higher the mixing height. Normally one expects the highest observed concentration (and correspondingly the largest the underprediction) to occur under the lowest mixing height. This result seems to be due to the fact that there are fewer cases in which the plume centerline exceeds the mixing height when the mixing height is quite high. There does not seem to be a significant correlation of the residuals with buoyancy flux for most models.

The PGT stability class correlates significantly positive with the residuals produced by most models. The positive correlation indicates that the larger underpredictions occur under the more stable conditions while any overpredictions occur during highly unstable conditions. This trend is consistent with the findings of earlier model performance studies, such as the EPA's rural model evaluation study, in which the largest overpredictions were also associated with unstable conditions and the largest underpredictions with stable conditions. However, unless additional data is supplied, a multivariate analysis will not make clear what effect stability-independent factors, such as interaction of the mixing height and plume centerline, have on the model performance statistics.

The scaled stability ratio U(10 m)/W_* seems to correlate significantly only with the CEQM residuals of SF_6 at Kincaid. The positive correlation suggests that the largest

TABLE 5-4. Correlation coefficients of predicted minus observed concentration residuals versus model input parameters for the 50 hours with the highest observed concentrations.

Study/Model	W_s (10 m)	Mixing Height	Buoyancy Flux	PGT Stability	V_{10}/W_* (10 m)	F_*
Clifty Creek/CRSTER	-0.08	-0.27	0.03	0.31	--	--
Kincaid SO_2/CRSTER	-0.00	-0.03	-0.13	0.04	0.05	-0.02
Kincaid SF_6/CRSTER	0.09	-0.20	-0.25	0.28	0.38	0.09
Bull Run SF_6/CRSTER	-0.33	0.03	0.12	0.61	-0.06	0.44
Maryland PPSP/CRSTER	-0.14	-0.25	0.03	0.22	0.05	0.10
Kincaid SO_2/TEM	-0.05	-0.06	-0.06	-0.09	-0.04	-0.07
Kincaid SF_6/TEM	0.05	-0.07	0.16	0.34	0.26	0.09
Bull Run SF_6/RTDM	-0.17	-0.13	0.03	0.32	0.13	0.47
Bull Run SF_6/SHORTZ	-0.23	0.30	-0.01	0.24	-0.20	0.28
Maryland PPSP/PPSP	-0.02	0.22	-0.04	0.09	-0.24	-0.27

underpredictions occur during the most stable conditions, which is consistent with findings when the residuals are stratified by the PGT stability class. The scaled buoyancy flux is a measure of the plume's tendency to rise to the top of the mixed layer and "bump" against the elevated inversion. The overpredictions for Bull Run seem to correlate significantly in a positive fashion with the scaled buoyancy flux. This suggests that although the plume may have risen to the top of the mixed layer, the observed final rise was not above the elevated inversion.

When interpreted without supporting statistics or graphics, correlation coefficients can often be misleading. In order to determine if such were the case, a number of scatter plots of relations between the residuals and model parameters were produced. A set of such plots for CEQM at Bull Run is presented as Figure 5-1. The scatter plots indicate the following:

> The PGT correlation of increasing underprediction with increasing stability is confirmed.
>
> The relations between mixing height, wind speed, buoyancy flux, and scaled stability ratio are inconclusive.
>
> The correlation of the scaled buoyancy flux with the residuals seems to be the result of a very few outlier values when plume penetration is very likely to have occurred. Removal of outliers results in an insignificant correlation.

We can see from Figure 5-1 that most of the scaled buoyancy flux values are less than 0.5 and occur for both over- and underpredictions.

The residuals from an application of CEQM to Kincaid, when SF_6 results were used for evaluation, exhibit no conclusive correlations with model parameters (Figure 5-2). The scaled buoyancy flux values for Kincaid are not significantly different from those for Bull Run. Again, there seems to be no systematic correlation of residuals with scaled buoyancy flux. The only appreciable tendency that appears in Figure 5-2 is that the largest overpredictions occur for large $U(10\ m)/W_*$ ratios (most stable condition) and the largest underpredictions occur during the smallest ratios (most unstable conditions).

In the plots of residuals from the CRSTER application at Clifty Creek (Figure 5-3), the correlations appear upon visual inspection of the scatter plots to be just marginal associations and not really significant. Only the PGT stability class exhibits any systematic relationship with the sign or magnitude of the residual. It can be seen, however, that this association has many exceptions.

The residuals for the PPSP model application in the Maryland PPSP study (Figure 5-4) generally exhibit a random relationship with the model input variables. The only exception is a slight tendency for overpredictions to occur for large F_*, which is possible because of the improved treatment of plume penetration cases. This random

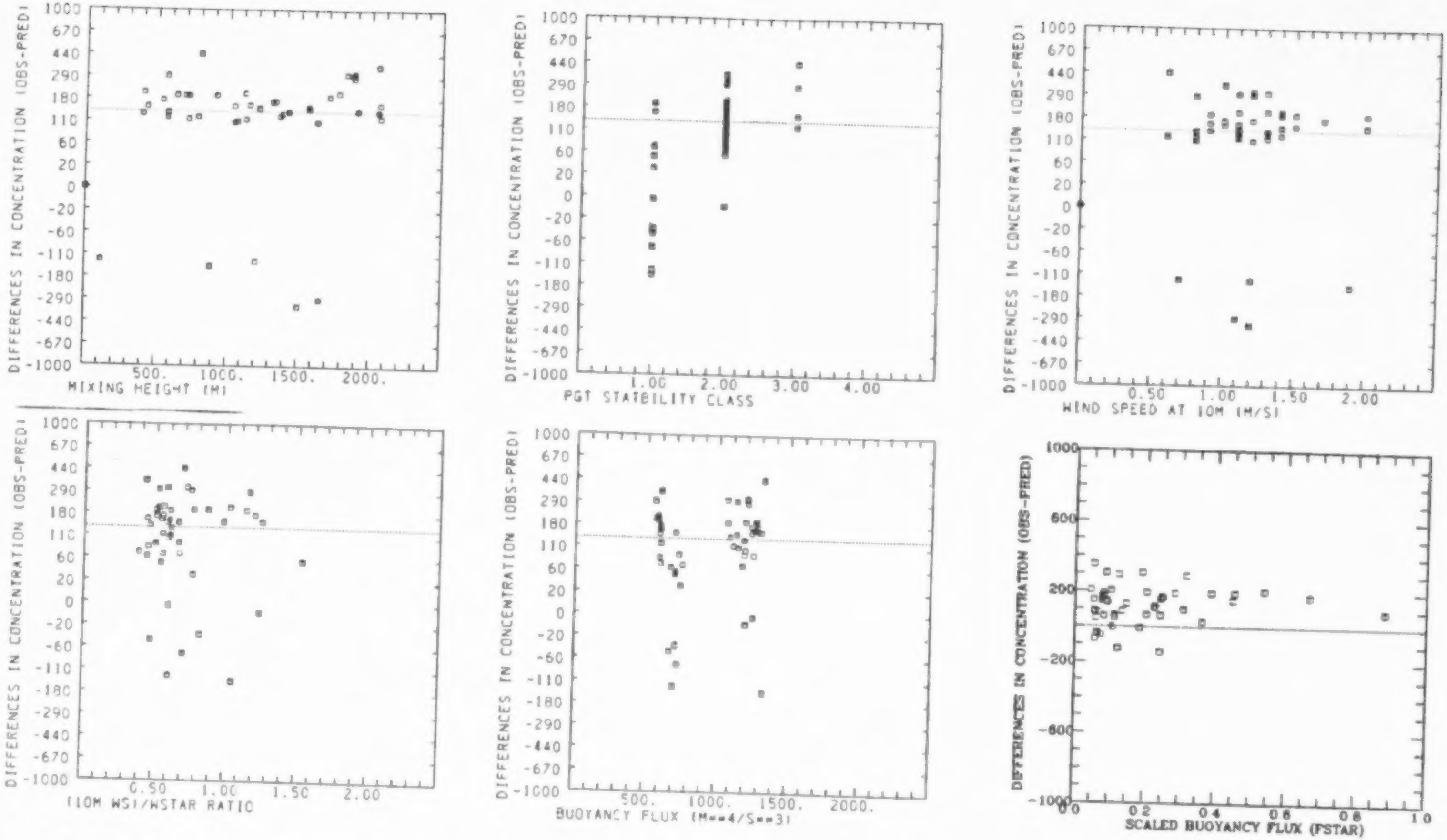


FIGURE 5-1. Scatter plots of residuals versus various model inputs and stability and scaling parameters for CEQM at Bull Run. The sample set is for the 50 highest observations. The units of the residuals are 10^{-9} s/m^3.

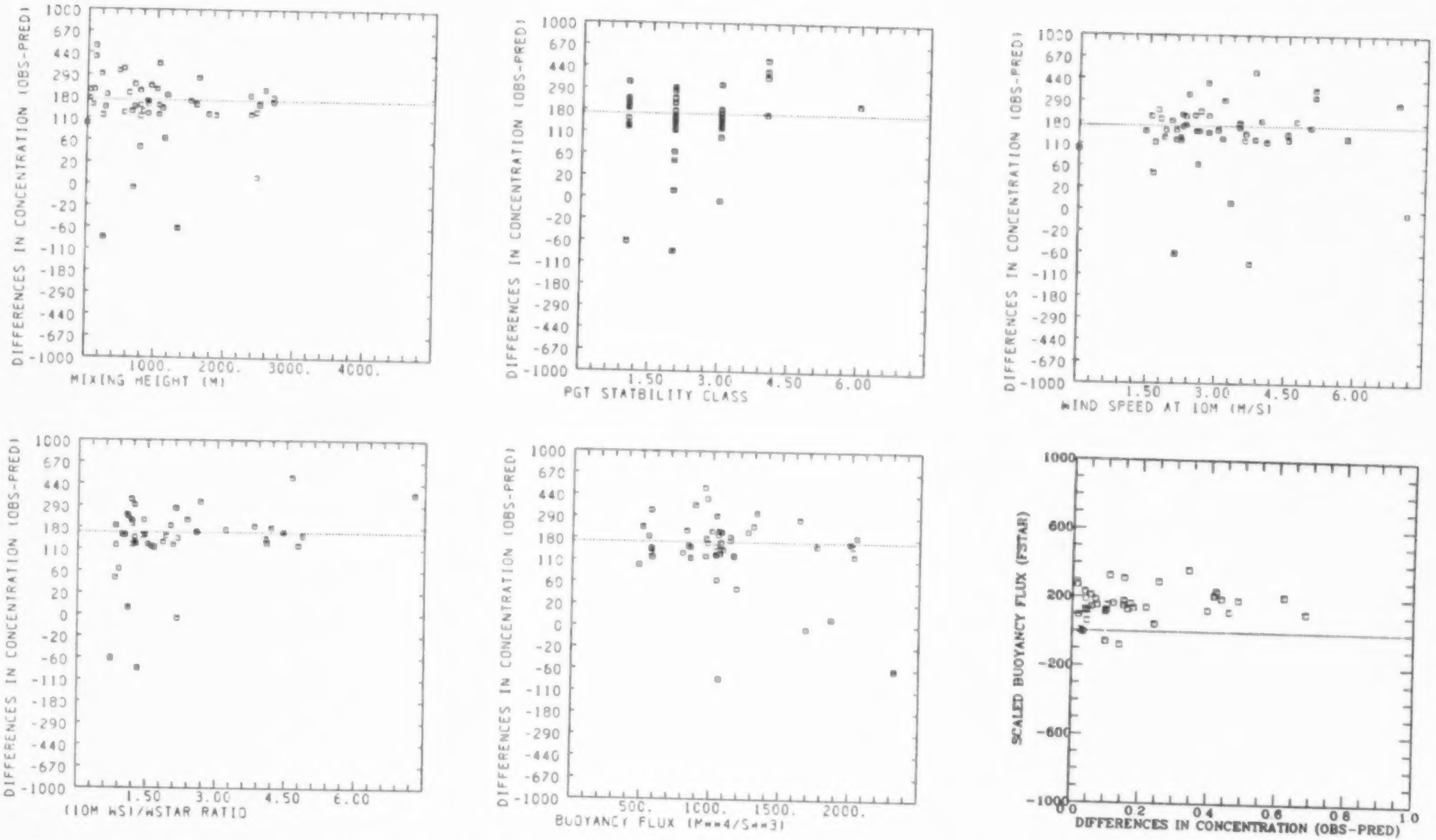


FIGURE 5-2. Scatter plots of residuals versus various model inputs and stability and scaling parameters for CEQM at Kincaid (SF_6). The sample set is for the 50 highest observations. The units of the residuals are 10^{-9} s/m^3.

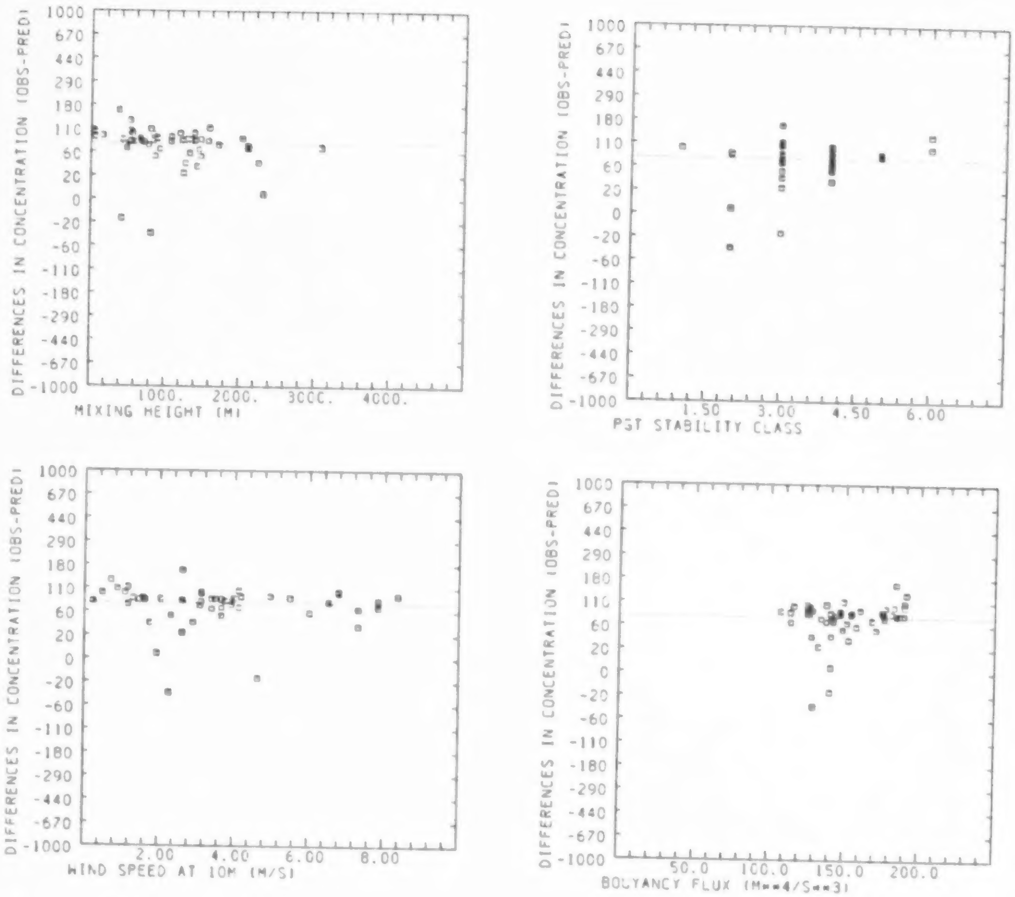


FIGURE 5-3. Scatter plots of residuals versus various model inputs and stability and scaling parameters for CRSTER at Clifty Creek. The sample set is for the 50 highest observations. The units of the residuals are 10^{-9} s/m^3.

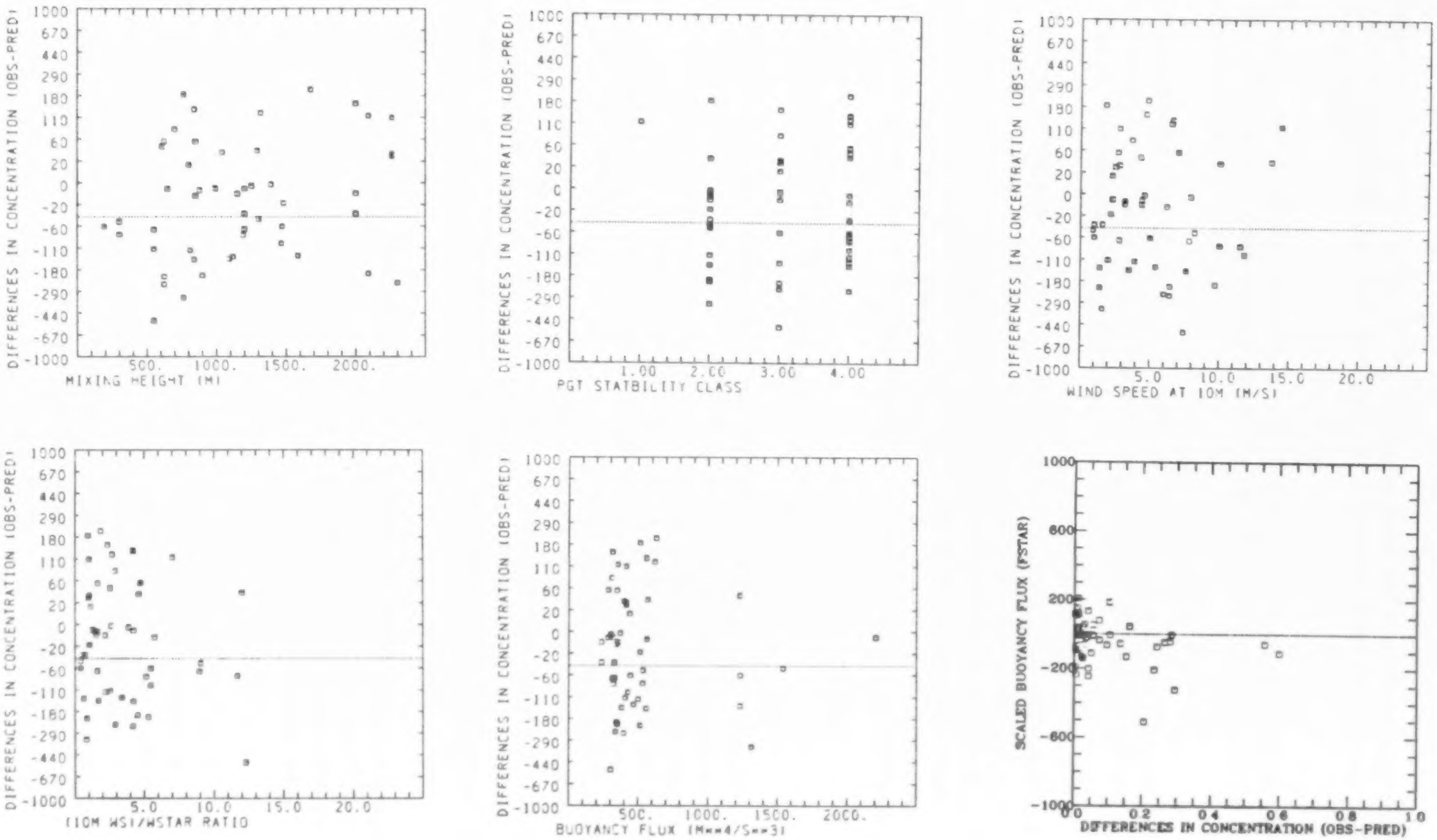


FIGURE 5-4. Scatter plots of residuals versus various model inputs and stability and scaling parameters for PPSP at the Maryland PPSP sites. The sample set is for the 50 highest observations. The units of the residuals are 10^{-9} s/m^3.

behavior also occurs in the CRSTER results from the Maryland study and in evaluations of the CEQM and TEM applications to Kincaid when SO_2 results were used for the evaluation.

In the scatter plots for residuals of the RTDM application to Bull Run the relationship with stability class clearly stands out, as does the relation with the scaled buoyancy flux (Figure 5-5).

One interesting feature seen in several of the scatter plots is that the buoyancy flux is bimodal with a clear split. However, because both groups are spread over a wide range of residuals, no correlation results.

The exploratory analysis relating the observation-prediction residuals to a single model input variable indicates that the association of PGT stability with the residuals is the most pervasive and strongest of all of the associations, although this association is not significant for several of the model performance evaluations. For the most part, however, no single factor stands out. As a result, we undertook a multivariate analysis of the 50 hours with the largest over- and underpredictions.

CLUSTER ANALYSIS OF THE LARGEST RESIDUALS

There are a limited number of meteorological conditions that result in large predicted or observed concentrations. The purpose of the cluster analysis is to group hours of the 50 largest magnitude residuals by the characteristics of the dispersion conditions that were occurring. It is not expected that all residuals can be identified with a statistically significant group. The reason for this lack of one-to-one matching is that the variables used to do the clustering (wind speed, mixing height, stability class, and buoyancy flux) do not provide a complete description of how the observed concentrations occur. As a result, an extreme overprediction may be grouped with a small overprediction or even an underprediction due to a lack of a discriminatory variable. Such groups become homogeneous and cannot be related to any specific set of dispersion conditions.

For those cases that make up significant clusters, the mean values and ranges of the variables producing the clustering can in principle be related to specific dispersion conditions, which in turn can be related to specific types of plume behavior. In our analysis we will attempt to produce significant clusters that can be used in the model performance synthesis.

A multivariate discriminatory analysis was conducted using the FASTCLUS procedure from SAS (SAS, 1985). This procedure uses the k-means algorithm, as described by Anderberg (1973). The procedure was set so that a maximum number of four clusters would be formed; this maximum was selected on the basis of how many dispersion regimes might reasonably be expected to exist. We lumped the highest over- and underpredictions into one set. The idea behind this was to see if there is a

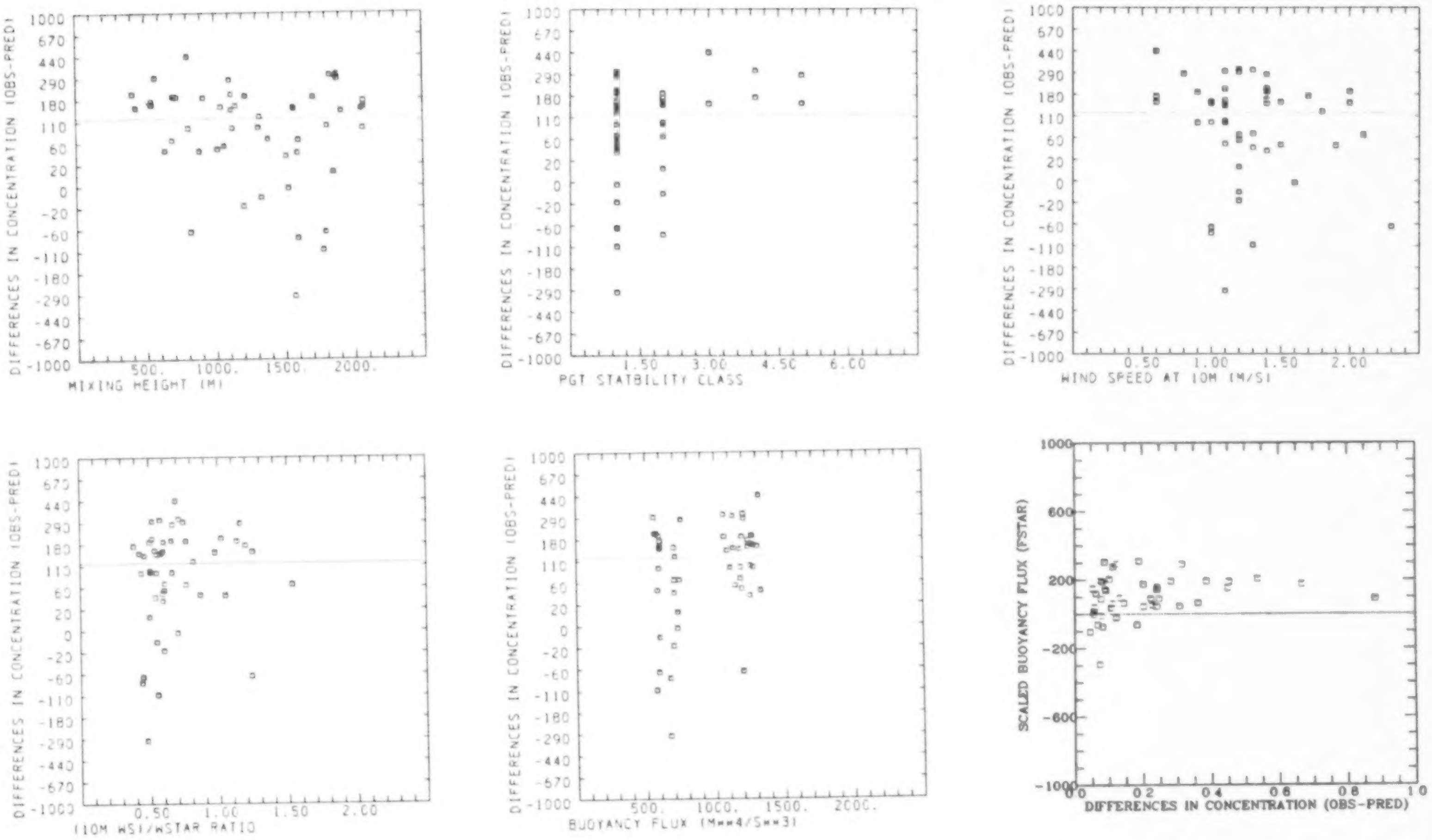


FIGURE 5-5. Scatter plots of residuals (observation minus prediction) of the 50 highest concentrations versus model inputs and stability and scaling parameters for RTDM at Bull Run. The units of the residuals are 10^{-9} s/m^3.

significant difference between the meteorology and buoyancy fluxes for under- and overpredictions or whether or not the underprediction cases "bleed" into the overprediction clusters. If the latter case holds, then only general and limited conclusions can be drawn about model performance.

The data were standardized before clustering in order to avoid the effects of differing means and variances. This leads to some difficulty in interpretation, but armed with the actual means and standard deviations one can recover actual cluster means. The clustering was done on two data sets. In the first, only the major model parameters (wind speed at 10 m, mixing height, PGT stability class, and buoyancy flux) are used in conjunction with the residuals. A second data set uses the mixing height, the stability ratio $U(10\ m)/W_*$, and the scaled buoyancy. This clustering is biased by the effects of interactions between variables because of the definitions of there variables. These sets of variables were used primarily because previous analyses (e.g., Pierce, 1988) also made use of these parameters.

At this point we are merely trying to identify general performance for each cluster rather than trying to pinpoint a dominant variable. Consequently, a hierarchical joining algorithm was not used to generate a cluster "tree". Because not all of the results of the cluster analyses can be presented, and in any case similar results were obtained from several models and data sets, a single example is presented here.

The clustering algorithm results for CEQM at Bull Run are summarized in Table 5-5. This table shows that overpredictions are grouped as cluster 4 with the following characteristics:

Highly unstable (Class A-B) conditions

Generally low wind speeds that are nevertheless significantly higher than those associated with underpredictions

Moderate mixing heights and buoyancy fluxes

The very largest underpredictions (cluster 2) are produced by nearly neutral conditions with low wind speed and higher than average buoyancy fluxes. A second group of underpredictions (cluster 1) is produced when mixing heights are high, wind speeds are low, the PGT stability class is approximately Class B, and the buoyancy fluxes are lower than average. A third group (cluster 3) has low mixing heights, an average stability that is approximately PGT class B, higher than average buoyancy fluxes, and generally low wind speeds. It seems that, with the exception of cluster 1, the underpredictions occur as the result of an "all-or-nothing" situation when the predicted final rise is above the mixed layer and the model predicts zero concentrations at the ground.

Table 5-6 indicates that most CEQM overpredictions at Bull Run occur under greater stability and lower scaled buoyancy flux. Underpredictions generally occur during

TABLE 5-5. Cluster standardized means from the 50 hours with the largest magnitude residuals for CEQM at Bull Run using routine model input data. The information in parentheses represents an interpretation of the mean value.

Variable	$\overline{X}$	σ	Cluster Number 1 (N = 12)	2 (N = 3)	3 (N = 24)	4 (N =11)
Residual	112	165	0.27 (U)	1.83 (U)	0.24 (U)	-1.32 (O)
PGT stability class	1.86	0.57	0.42 (B)	1.51 (C-D)	0.18 (B)	-1.27 (A)
Mixing height	1198	545	1.19 (H)	0.44 (M)	-0.61 (L)	-0.09 (M)
Buoyancy flux	1014	288	-0.54 (L)	1.02 (H)	0.28 (M)	-0.30 (M)
U (10 m)	1.18	0.32	-0.68 (L)	-1.09 (L)	0.10 (L)	0.83 (L)

TABLE 5-6. Cluster standardized means from the 50 hours with the largest magnitude residuals for CEQM at Bull Run using convective scaling variables. The information in parentheses represents an interpretation of the mean value.

Variable	$\overline{X}$	σ	Cluster Number 1 (N = 26)	2 (N = 2)	3 (N = 2)	4 (N = 20)
Residual	112	165	0.20 (U)	1.47 (U)	0.73 (U)	-0.48 (O)
Mixing height	1198	545	0.60 (H)	-1.27 (L)	-0.29 (M)	-0.63 (M)
U (10 m)/W_*	--	--	-0.64	-0.02	1.94	0.61
Scaled buoyancy flux	--	--	0.32	5.68	2.04	-0.09

U = underprediction
O = overprediction
A-F = PGT stability class
L = low
M = moderate
H = high

$\overline{X}_c$ = cluster mean (standardized)
$\overline{X}$ = variable mean
σ = variable standard deviation

highly unstable conditions and with higher scaled buoyancy fluxes. The very largest underpredictions occur when the scaled buoyancy flux is very large (clusters 2 and 3) and thus may have plume centerline elevations higher than the mixing height. Cluster 3 occurs under nearly neutral conditions. For most models and sites the use of scaled stability and buoyancy produces poorer discrimination of groups of residuals. An examination of individual hours present in each cluster indicates that at Bull Run some of the small underprediction hours "bleed" into the overprediction category.

The largest CEQM overpredictions at Kincaid occur under high wind speed and low buoyancy flux conditions, which results in plume rises that may be too close to the ground and which at any rate would not result in the plume centerline being above the elevated inversion (Table 5-7). Other underpredictions occur under convectively unstable conditions with large buoyancy flux and probably represent hours where the plume undergoes either looping or "bumping" behavior.

The CEQM underpredictions occur primarily under PGT stability class B conditions, moderate mixing hights, and when lower than average buoyancy fluxes and wind speeds occur. Interestingly enough, cluster 1 in Table 5-7 shows a few of the larger underpredictions ocurring under neutral conditions. Similar clustering also shows up for all the model performance data sets at Kincaid. The major difference between the SF_6 and SO_2 clustering at Kincaid is that high winds occurs with low buoyancy for the SO_2 clustering. This type of clustering suggests that conditions leading to low plume rise and neutral coning plume behavior are typical for CEQM underpredictions of SO_2 at Kincaid.

CRSTER underpredictions at Clifty Creek occur under nearly neutral stability conditions with low mixing heights (Tale 5-8). One cluster (cluster 2) has low buoyancy fluxes and high wind speeds, while another (cluster 4) has high buoyancy and low wind speeds. In the latter case the final plume rise is high; as a result, the model often predicts zero concentrations at the ground because the final plume centerline is above the mixing height. In the former case the plume rise is low, but under nearly neutral conditions the dispersion algorithm may be predicting a wider or higher plume than actually occurs.

The overpredictions made by CRSTER at Clifty Creek fall into two groups. According to Table 5-8 most of the overpredictions occur under unstable conditions, PGT class B, and with moderate mixing heights, buoyancy flux, and wind speeds. This suggests that the underpredictions are occurring when the plume is in a convective dispersion regime and does not rise to the top of the mixed layer. The second group of overpredictions (cluster 3) occur under unstable conditions as well, except that the winds and buoyancy flux are low and the mixing height is high. Thus, as in the first group of overpredictions, the plume does not rise above the elevated inversion, but spreads rather rapidly. In this case the overprediction may be partly due to the possibility that although the plume is following the terrain more, the CRSTER algorithm assumes a lower plume elevation above the ground than actually occurs. Of course

TABLE 5-7. Cluster standardized means from the 50 hours with the largest magnitude residuals for CEQM at Kincaid (SF_6). The information in parentheses represents an interpretation of the mean value.

Variable	$\bar{X}$	σ	Cluster Number			
			1 (N = 6)	2 (N = 18)	3 (N =8)	4 (N = 18)
Residual	15	264	1.10 (U)	0.65 (U)	-1.00 (O)	-0.57 (O)
PGT stability class	2.34	1.17	2.00 (D)	-0.32 (B)	0.15 (B-C)	-0.42 (B)
Mixing height	885	839	-0.28 (L)	0.05 (M)	0.12 (M)	-0.02 (M)
Buoyancy flux	1248	54	-0.43 (M)	-0.50 (M)	-1.06 (L)	1.11 (H)
U (10 m)	3.44	1.21	-0.27 (M)	-0.42 (L)	1.30 (H)	-0.06 (M)

TABLE 5-8. Cluster standardized means from the 50 hours with the largest magnitude residuals for CRSTER at Clifty Creek. The information in parentheses represents an interpretation of the mean value.

Variable	$\bar{X}$	σ	Cluster Number			
			1 (N = 27)	2 (N = 4)	3 (N = 15)	4 (N = 4)
Residual	-80	80	-0.42 (O)	1.55 (U)	-0.33 (O)	2.54 (U)
PGT stability class	2.67	0.80	-0.36 (B)	1.05 (C-D)	0.01 (B-C)	1.36 (D)
Mixing height	1184	564	-0.32 (M)	-0.68 (L)	0.93 (H)	-0.69 (L)
Buoyancy flux	141	24	0.31 (M)	-0.81 (L)	-0.75 (L)	1.56 (H)
U (10 m)	2.69	1.14	-0.07 (M)	2.00 (H)	-0.92 (L)	-1.19 (L)

U = underprediction
O = overprediction
A-F = PGT stability class
$\bar{X}$ = variable mean
σ = variable standard deviation

L = low
M = moderate
H = high

part of the overprediction could also be caused by too great a rate of vertical dispersion at unstable conditions, like that occurring in the other studies.

Because the PPSP model has no zero underpredictions, we felt that it would be most possible in evaluations of this model for the model underpredictions to "bleed" extensively into that of the overpredictions. Table 5-9 confirms the suspicion, as we can see that cluster 1 contains a mixture of over- and underpredictions that cannot be separated out. The cluster combining both over- and underpredictions occurs during neutral conditions with typical (near average) values of mixing height, buoyancy flux, and wind speed.

The large overpredictions appear to be sortable into three groups (Table 5-9). Interestingly, the largest overpredictions occur under slightly unstable conditions with lower than average mixing heights. This suggests that perhaps the model is underestimating dispersion during plume trapping, resulting in a predicted plume that is too compact. The two other groups occur under unstable conditions. In cluster 3 the buoyancy flux is larger than average, suggesting that models may not be accounting for possible enhanced spread by plume "bumping" or mass lost to partial penetration, so the predicted plume is either too compact or else has too much mass.

For cluster 4 the buoyancy flux is low and the wind speeds are high, resulting in low rise. The model seems to be predicting too much dispersion for the low plume rise conditions and lower edge of the plume is reaching the ground.

A cluster analysis using scaling quantities was done on the PPSP model data set with the intent of trying to distinguish between underprediction dispersion conditions and overprediction conditions. However, we found that underpredictions still "bleed" into overpredictions (Table 5-9). A set of overpredictions with high buoyancy flux stands out, as does a cluster with high winds and limited plume rise. Most of the greatest overpredictions again occur under moderate conditions as was found by the cluster analysis using model input variables.

The clustering of residuals for RTDM at Bull Run was examined (Table 5-10). Overpredictions (cluster 4) occur most often under highly unstable conditions with low buoyancy fluxes such that the plume lies below the elevated inversion, which is also lower than average. As a result, the plume may be looping or bumping, two phenomena that lead to increased dispersion and which are not treated by the PG dispersion curves.

The underpredictions occur under both highly unstable and neutral conditions. Surprisingly, the greatest underprediction occurs under unstable conditions (clusters 1 and 3). There is nothing outstanding about either group that would suggest that the underpredictions occur because of the "all-or-nothing" plume penetration algorithm used in the models. The tendency to underpredict under stable-neutral conditions is similar to that discovered for other models and other sites. At Bull Run the winds are generally low and the mixing heights are high, leading to plumes that rise above

TABLE 5-9. Cluster standardized means from the 50 hours with the largest magnitude residuals for the PPSP model during the Maryland PPSP. The information in parentheses represents an interpretation of the mean value.

Variable	$\overline{X}$	σ	Cluster Number			
			1 (N = 23)	2 (N = 16)	3 (N = 7)	4 (N = 4)
Residual	-91	150	0.54 (O/U)	-0.70 (O)	-0.15 (O)	-0.06 (O)
PGT stability class	3.4	0.81	0.42 (D)	-0.03 (C)	-0.32 (B)	-1.73 (A-B)
Mixing height	1188	582	0.14 (M)	-0.76 (L)	0.24 (M)	1.79 (H)
Buoyancy flux	568	326	-0.34 (M)	-0.30 (M)	2.18 (H)	-0.69 (L)
U (10 m)	4.76	2.49	0.13 (M)	-0.34 (M)	-0.27 (M)	1.06 (H)
			1 (N = 6)	2 (N = 3)	3 (N = 7)	4 (N = 34)
Residual	-91	150	-1.51 (O)	-0.83 (O)	-0.13 (O)	0.37 (O/U)
Mixing height	1188	582	0.62 (H)	-1.24 (L)	-0.99 (L)	-0.21 (M)
U (10 m)/W_*	--	--	-0.21 (M)	2.77 (H)	-0.55 (L)	-0.09 (M)
Scaled buoyancy flux	--	--	-0.28 (M)	0.35 (M)	2.05 (H)	-0.40 (L)

U = underprediction
O = overprediction
A-F = PGT stability class
L = low
M = moderate
H = high
$\overline{X}$ = variable mean
σ = variable standard deviation

TABLE 5-10. Cluster standardized means from the 50 hours with the largest magnitude residuals for RTDM at Bull Run. The information in parentheses represents an interpretation of the mean value.

			Cluster Number			
Variable	$\overline{X}$	σ	1 (N = 26)	2 (N = 8)	3 (N = 12)	4 (N = 4)
Residual	138	124	0.11 (U)	0.29 (U)	0.47 (U)	-2.73 (O)
PGT stability class	1.8	1.12	-0.20 (A-B)	1.62 (D)	-0.49 (A)	-0.49 (A)
Mixing height	1189	569	-0.22 (M)	1.04 (H)	-0.10 (M)	-0.37 (L)
Buoyancy flux	1006	284	0.67 (M)	-0.54 (M)	-0.66 (M)	-1.31 (L)
U (10 m)	1.22	0.39	-0.02 (M)	-1.11 (L)	0.93 (M)	-0.43 (M)

U = underprediction
O = overprediction
A-F = PGT stability class
L = low
M = moderate
H = high
$\overline{X}$ = variable mean
σ = variable standard deviation

the elevated inversion. Other cases could be due to wind speeds aloft that are higher than observed so that the predicted plume rise is too high.

DISCRIMINANT ANALYSIS OF THE RESIDUALS

In Tables 5-5 through 5-9 it is not clear which variables may be most important in establishing the clustering centroid. In order to determine the combination of variables that have the highest possible multiple correlation within cluster groups, we used the SAS CANDISC procedure (SAS, 1985). Canonical variables are linear combinations of the original variables (e.g., $CV = \sum \alpha_i x_i$, where x_i is the ith original variable and α_i is a weight), much like principal components, or factors. Canonical variables are uncorrelated and their weights represent how much each of the original variables contributes to the canonical variable; this procedure is similar to factor analysis. (For more information on the canonical discrimination procedure the reader is referred to pages 156-157 of the SAS user's manual, 1985.) Use of canonical variables offers a method of determining which variables are most important in determining which hours are joined to a cluster. Rather than examining individual ratios of between-group variance to within-group variance for each variable, we can examine the clustering in terms of only the two dominant canonical variables.

The canonical variables allow us to examine fewer scatterplots of the clusters since the first two variables define that slice of N-dimensional variable space that produces the most organized (correlated) scatter. If by chance a particular variable dominates the weights making up the canonical variable, the variations in that variable can be thought of as being closely related to the variations in the canonical variable.

We computed canonical variables for the clustering done for the models and sites. To some degree these variables can be treated as principal components to which they are closely related. We inspected the coefficients for dominant variables. Table 5-11 presents the coefficients for each clustering exercise. The dominant variables are printed in bold. It can be seen from this table that the residuals and the buoyancy flux dominate most often. It is these two variables that are usually most important in determining cluster formation.

The scatter plots illustrate how organized the cluster "boundaries" are. If the clusters are distorted and interlaced with other clusters, the within-group variance is comparable to the between-group variance and the clustering is not very effective. In order to examine how "robust" the clustering is, we visually inspected plots of the cluster number that are projected on the plane defined by the first two canonical variables.

Figure 5-6 shows that the clusters are generally grouped by region and hours from one cluster do not mix much with other clusters. Only cluster 2 shows a wild degree

TABLE 5-11. A summary of the canonical coefficients (weights) for the two canonical variables that explain most of the variance within the clusters. Bold numbers indicate dominant variables.

Model/Study	Residual		PGT Stability Class		Mixing Height		Buoyancy Flux		W_s (10 m)	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
CEQM/Bull Run	1.23	-0.27	0.83	-0.14	0.42	1.26	0.34	-0.30	-0.58	-0.09
RTDM/Bull Run	1.42	1.11	-0.98	0.96	-0.40	0.51	0.29	-0.81	-0.68	0.37
SHORTZ/Bull Run	0.14	-0.01	0.71	-0.31	1.36	0.47	-0.24	-0.68	0.00	1.11
CRSTER/Clifty Creek	2.63	0.32	0.14	-0.03	0.19	0.68	-0.12	-1.19	0.26	0.76
PPSP/Maryland PPSP	-0.33	0.45	-0.47	-0.51	0.14	1.11	2.28	-0.14	0.20	0.34
CRSTER/Maryland PPSP	0.42	-0.76	-0.63	0.90	0.67	-0.90	1.62	1.30	-0.68	0.62
CEQM/Kincaid SF_6	1.13	0.87	1.15	0.27	-0.10	-0.17	-1.92	0.84	0.00	0.70
CEQM/Kincaid SO_2	0.48	1.48	0.42	-0.65	-0.52	0.80	-0.42	0.66	1.29	0.17
TEM/Kincaid SF_6	0.45	0.66	0.87	0.98	-0.22	0.16	-1.65	0.87	0.32	0.02
TEM/Kincaid SO_2	0.10	0.63	1.03	-0.56	0.00	0.43	0.53	0.58	0.87	1.18

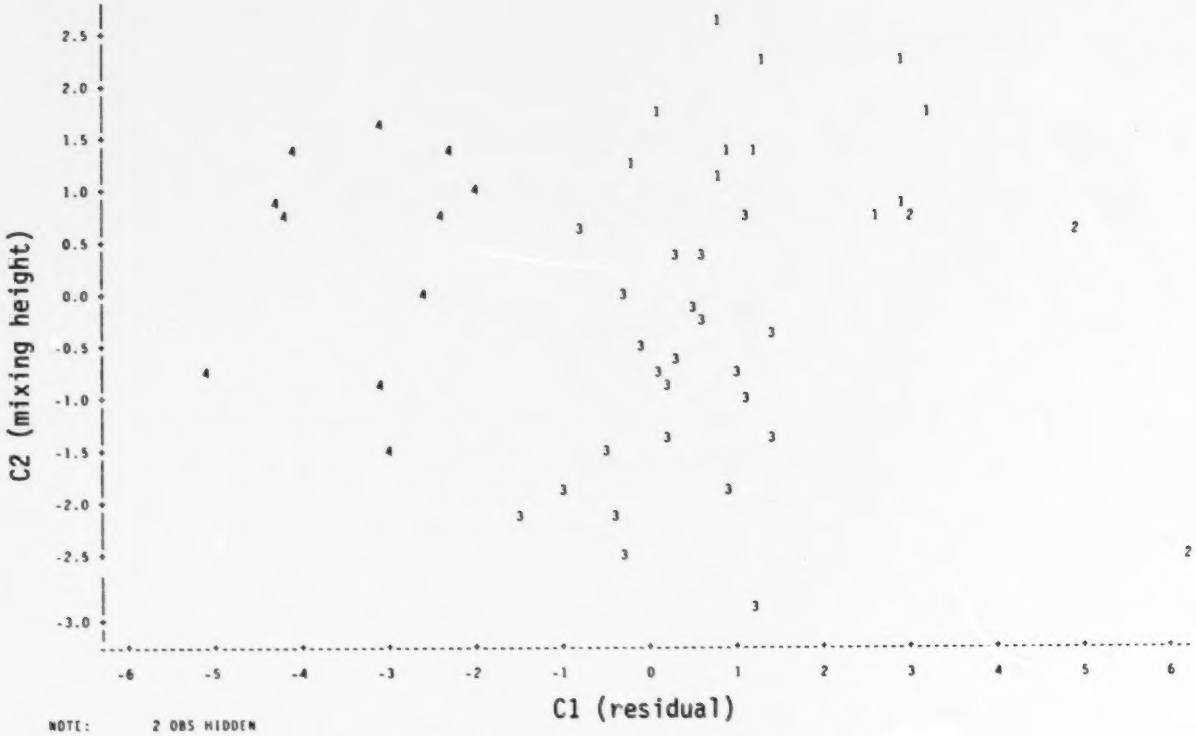


FIGURE 5-6. A scatter plot of case cluster number versus the first two canonical variables. The dominant original variable is indicated in parentheses. The plot is for CEQM at Bull Run.

of scatter. Cluster 2 could be justifiably split again, but its importance is limited since its sample size is very small. Figure 5-7 shows an example of clusters that have poor discriminatory capability, such as occurred with the PPSP model. Here we see that the cluster numbers when projected onto the plane defined by the two dominant canonical variables are not grouped into well-defined clusters. Fortunately, most of the scatter plots for the model evaluation studies showed a significant degree of regional compartmentalization. This result indicates that the cluster groupings from the previous analysis are generally significant and do not arise from an arbitrary procedure.

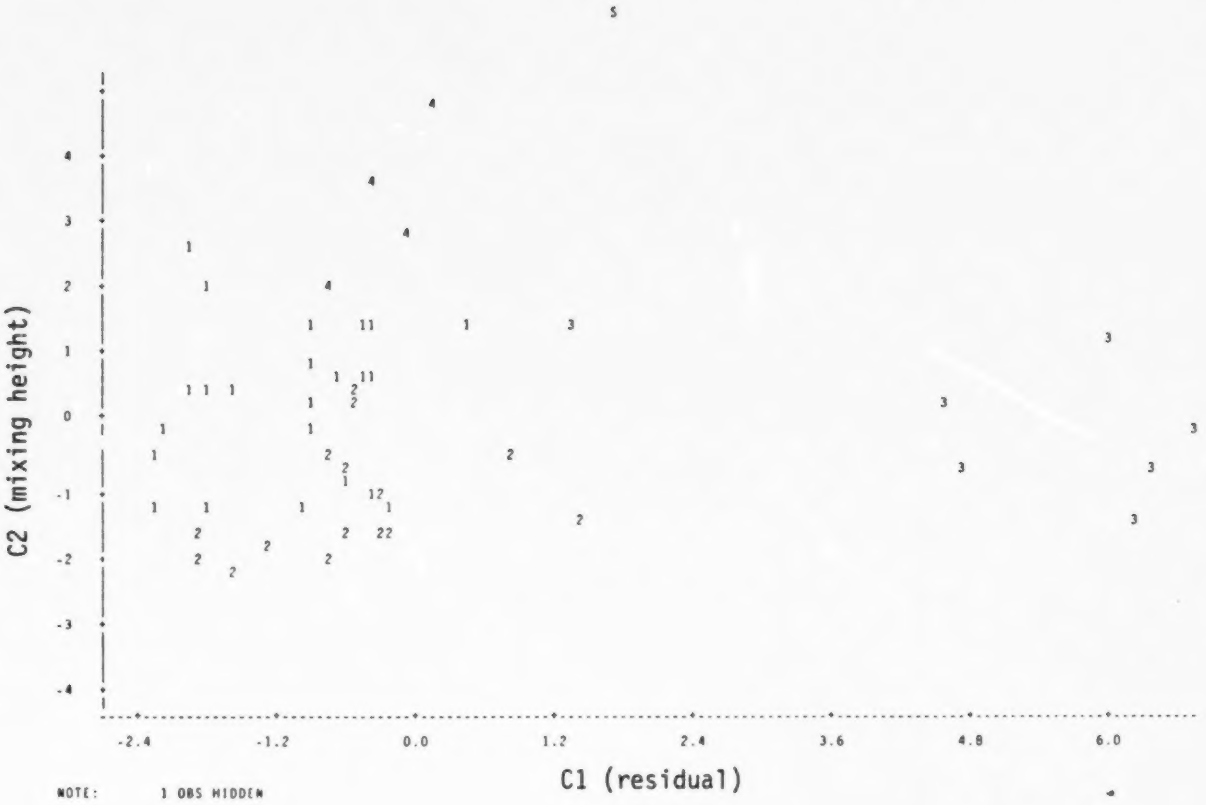


FIGURE 5-7. A scatter plot of case cluster number versus the first two canonical variables. The dominant original variable is indicated in parentheses. The plot is for PPSP at the Maryland PPSP sites.

6 SYNTHESIS, SUMMARY AND CONCLUSIONS

We have conducted an analysis of the results of previous dispersion model performance evaluation studies to ascertain the suitability of currently used models for the prediction of hourly average concentrations. This investigation was motivated by the potential promulgation of a statistical (expected exceedance) 1-hour SO_2 NAAQS. Implementation of such a standard would rely largely on the use of dispersion models that are capable of accurately predicting the whole upper tail of the distribution of hourly average SO_2 concentrations, not just the highest or second highest concentrations.

Our analysis was confined to the study of model performance for elevated, buoyant point sources of the type typically associated with fossil fuel-fired power plants. Such sources represent one of the major areas of model application for the prediction of SO_2 air quality impacts. We conducted analyses designed to address the following issues:

> How well do models predict the 1, 2, ..., N highest concentrations irrespective of time or location matching?
>
> How well do models predict the 1, 2 ,..., N highest observed concentrations when observations and predictions are matched in time but not location?
>
> How well do models predict concentrations occurring under specific meteorological conditions? More specifically, what meteorological conditions produce the highest observed concentrations? What conditions are associated with the largest discrepencies between observed and predicted concentrations?

To answer the above questions, we reviewed the published results of previous model performance studies and, whenever possible, obtained the aerometric and meteorological data from those studies for further analysis. Our results are summarized in the following sections.

MODEL PERFORMANCE: ABILITY TO PREDICT PEAK CONCENTRATIONS

Short term (1- to 24-hour average) ambient air quality standards are designed to control the magnitude and frequency of occurrence of peak ground-level concentrations. Demonstrating that the emissions from a major source do not violate such standards often involves the use of a dispersion model to predict the highest concentrations occurring over a one- to five-year period in the vicinity of the source. A model's ability to accurately and precisely predict peak concentrations is therefore very important. However, a model need not necessarily predict the exact time and location of the peak values as this information is not usually of direct relevance to the attainment demonstration. On the other hand, examining a model's ability to reproduce the N highest observed concentrations during the hours in which they occur (i.e., matching concentrations in time) provides insight into the types of meteorological conditions under which models do especially well or poorly and the reasons for such performance. This information, in turn, can be used as a guide to improving models and to the level of performance that can be anticipated from a given model under a given situation.

Concentrations Unmatched in Time or Location

We examined the performance of models in predicting the N highest concentrations (typically around 25) regardless of their time or location of occurrence. In some cases, only comparisons of the highest and second-highest observed and predicted concentrations were available. We caution the reader that it is difficult to draw generalizations from such comparisons because the highest and second-highest concentrations usually have a large degree of variability.

Table 6-1 summarizes model performance in predicting the N highest observed concentrations when predictions and observations are unmatched in time or location. With the exception of the Maryland Power Plant Siting study (for which a value of N = 50 was used), the CRSTER and equivalent models tend to overpredict slightly the peak concentrations (N less than or equal to 25) or produce no significant bias. TEM, on the other hand, underpredicts at some sites and produces no significant bias at others. The only other models for which evaluation results are available from more than one site are PPSP and MPSDM. PPSP overpredicts consistently at all sites, in some cases by more than a factor of two. MPSDM also produced significant overpredictions.

Model Performance for Hourly Averages Compared with Performance for 3- and 24-Hour Averages

Would implementation of a 1-hour average NAAQS for SO_2 result in significantly greater modeling uncertainty than currently exists in demonstrating attainment of the 3- and 24-hour NAAQS? To answer this question, we reviewed an analysis of the

TABLE 6-1. Summary of model performance for peak concentrations unpaired in time or location. (N/A = results not available; NS = not significant).

	CRSTER/MPTER/CEQM			TEM			RTDM			SHORTZ					
	Avg.	High	Second High	Avg.	High	Second High	Avg.	High	Second High	Avg.	High	Second High	PPSP	RAMR	MPSDM
EPRI PMV															
Kincaid (SO_2)[5]								N/A			N/A		N/A	N/A	N/A
Bias[1]	NS	(Over)	(Over)	NS	(Under)	(Under)									
Magnitude (%)[2]	3	-24	-18	14	28	31									
Basis[3]	N = 25	R = 1	R = 2	N = 25	R = 1	R = 2									
No. of Hours[4]	1546	1546	1546	1702	1702	1702									
Kincaid (SF_6)[5]								N/A			N/A		N/A	N/A	N/A
Bias	NS	(Over)	(Over)	NS	(Under)	(Under)									
Magnitude (%)	-11	-45	-39	25.	17	19									
Basis	N = 25	R = 1	R = 2	N = 25	R = 1	R = 2									
No. of Hours	128	128	128	218	218	218									
Bull Run[6]					N/A								N/A	N/A	N/A
Bias	NS	(Under)	(Over)				NS	(Over)	(Over)	(Under)	(Under)	(Under)			
Magnitude (%)	12	2	-16				12	-9	-30	45	22	19			
Basis	N = 25	R = 1	R = 2				N = 25	R = 1	R = 2	N = 25	R = 1	R = 1			
No. of Hours	255	255	255				255	255	255	255	255	255			
EPA Power Plant Study															
Stuart, Philo, Canal, Muskingum[7]					N/A			N/A			N/A		N/A	N/A	N/A
Bias		(Over)	(Over)												
Magnitude (%)		--	--												
Basis		R = 1	R = 2												
No. of Hours		--	--												
Clifty Creek		N/A			N/A			N/A			N/A		N/A	N/A	N/A
Bias															
Magnitude (%)															
Basis															
No. of Hours															
Maryland Power Plant Siting Study[8]					N/A			N/A			N/A			N/A	N/A
Bias	(Under)												(Over)		
Magnitude (%)	--												--		
Basis	N = 50												N = 50		
No. of Hours	145												145		

continued

TABLE 6-1. Continued

	CRSTER/MPTER/CEQM			TEM			RTDM			SHORTZ					
	Avg.	High	Second High	Avg.	High	Second High	Avg.	High	Second High	Avg.	High	Second High	PPSP	RAMR	MPSDM
API Tracer Study[9]				N/A			N/A			N/A					
Bias	(Over)												N/A	(Over)	N/A
Magnitude (%)	--													--	
Basis	N = 10													N = 10	
No. of Hours	217													217	
EPA Rural Study[10]															
Clifty Creek							N/A			N/A					
Bias	(Over)			NS									(Over)	N/A	(Over)
Magnitude (%)	-33			-8									-434		-42
Basis	N = 25			N = 25									N = 25		N = 25
No. of Hours	7703			7618									9138		7830
Kincaid							N/A			N/A					
Bias	NS			Under									(Over)	N/A	(Over)
Magnitude (%)	5			10									-442		-55
Basis	N = 25			N = 25									N = 25		N = 25
No. of Hours	5172			6983									7461		6009
Muskingum							N/A			N/A					
Bias	(Over)			NS									(Over)	N/A	(Over)
Magnitude (%)	-74			-9									-217		-55
Basis	N = 25			N = 25									N = 25		N = 25
No. of Hours	2977			2977									2977		2977
Paradise							N/A			N/A					
Bias	NS			NS									(Over)	N/A	(Over)
Magnitude (%)	-9			5									-246		-82
Basis	N = 25			N = 25									N = 25		N = 25
No. of Hours	5499			5499									5499		5499
EPA Urban Study (St. Louis)[11]	N/A														
Bias				Over			N/A			N/A			N/A	Under[12]	N/A
Magnitude (%)				107										16	
Basis				N = 25										N = 25	
No. of Hours				4758										4758	

TABLE 6-1. Concluded

Notes

[1] Average residual over N highest observed and predicted concentrations or residual of R^{th} highest concentration unmatched in time or location: Over = overprediction; Under = underprediction; ( ) indicates bias is not significant at the 90 percent level or significance is unknown.

[2] Ratio of average residual (observed-predicted) over N highest or residual of R^{th} highest concentration unmatched in time or location to corresponding average or R^{th} highest observed concentration.

[3] N indicates bias is based on average residual over the N highest observed and predicted. R indicates bias is based on residual of the R^{th} highest observed and predicted.

[4] Number of hours in data set from which N or R highest determined.

[5] Indicated bias found using routine on-site meteorological data (level Ib data).

[6] Only SF_6 data is available at this site.

[7] Bias was consistent from site to site. Confidence intervals unavailable.

[8] From Figure 1 of Weil and Brower (1984). Confidence intervals unavailable.

[9] Based on Goodyear site results only (Londergan et al., 1980; Table 5-3).

[10] Results from Appendix A in Londergan et al., 1983. A 95 percent confidence interval was used.

[11] Compiled from Table 2-10 in this report.

[12] Results shown are for RAM (not RAMR).

variation in model performance with averaging time (Hayes et al., 1984). This study showed that the ratio of the average of the 25 highest predictions to the average of the 25 highest observations (unmatched in time or location) decreases with increasing averaging time for all models included in the EPA Rural Model Evaluation Study (Figure 6-1), the CEQM model at EPRI PMV Kincaid, and for RAM (but not TEM) in the EPA Urban Model Evaluation Study (not shown). Results for ratios of the highest and second-highest concentrations were mixed (Hayes et al., 1984; Figures 4-17 through 4-21). Thus, models that overpredict peak 3- and 24-hour averages tend to overpredict peak hourly averages by a greater amount, while models that underpredict peak 3- and 24-hour averages may overpredict peak hourly averages. This suggests that, at least as far as bias goes, shifting emphasis from 3- and 24-hour averages to 1-hour averages will result in more conservative predictions of peak concentrations.

The study also examined the correlation between observations and predictions paired in time but not location as a function of averaging time at the EPRI PMV Kincaid site using the CEQM model. The results (reproduced here in Figure 6-2) indicate that correlations for 3-hour averages are only slightly better than those for hourly averages, while the correlation for 24-hour averages is not significantly different from that of hourly averages. Thus, the model's skill in predicting hourly averages does not appear to be very much worse than it is for 3- and 24-hour averages.

Concentrations Matched in Time

A model's ability to predict peak concentrations can also be examined by comparing the N highest observed concentrations with the highest concentrations predicted during the same hours (i.e., compare concentrations matched in time but not location for the hours with the N highest observed concentrations). Such a comparison evaluates the model's ability to reproduce high observed concentrations under the specific meteorological conditions that cause them (exclusive of wind direction).

As has been demonstrated by the EPRI PMV and other studies, correlations between time-matched observed and predicted concentrations are generally very low or not significantly different from zero (see, for example, Figure 2-1 of this report as well as Figure 6-1 and Table 6-1 of both Reynolds et al., 1984 and 1985)*. Correla-

---

* Two major exceptions to this should be noted: The PPSP model produced predictions that correlated well with observations in the Maryland Power Plant Siting Study (correlation coefficient = 0.76; see Figure 2-1) and RTDM predictions correlated reasonably well with observations at the EPRI PMV Bull Run site when all hours with observed or predicted SF_6 concentrations greater than 2 ppt were considered (correlation coefficient = 0.46; see Reynolds et al., 1985, Table 6-1). One reason the PPSP model may have performed so well in the Maryland study is that the model was developed on the basis of that data set.

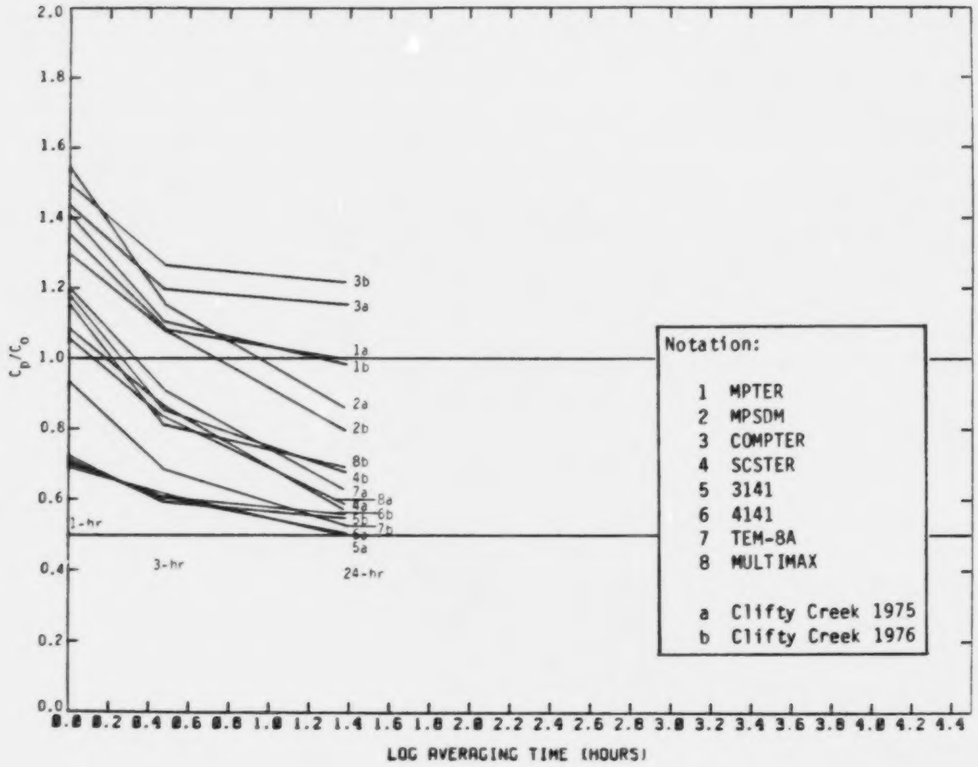


FIGURE 6-1. Dependence of model bias on averaging time in the EPA Rural Study (Clifty Creek 1975 and 1976 data). Ratio of means of highest 25 predicted and observed SO_2 values. (From Hayes et al., 1984).

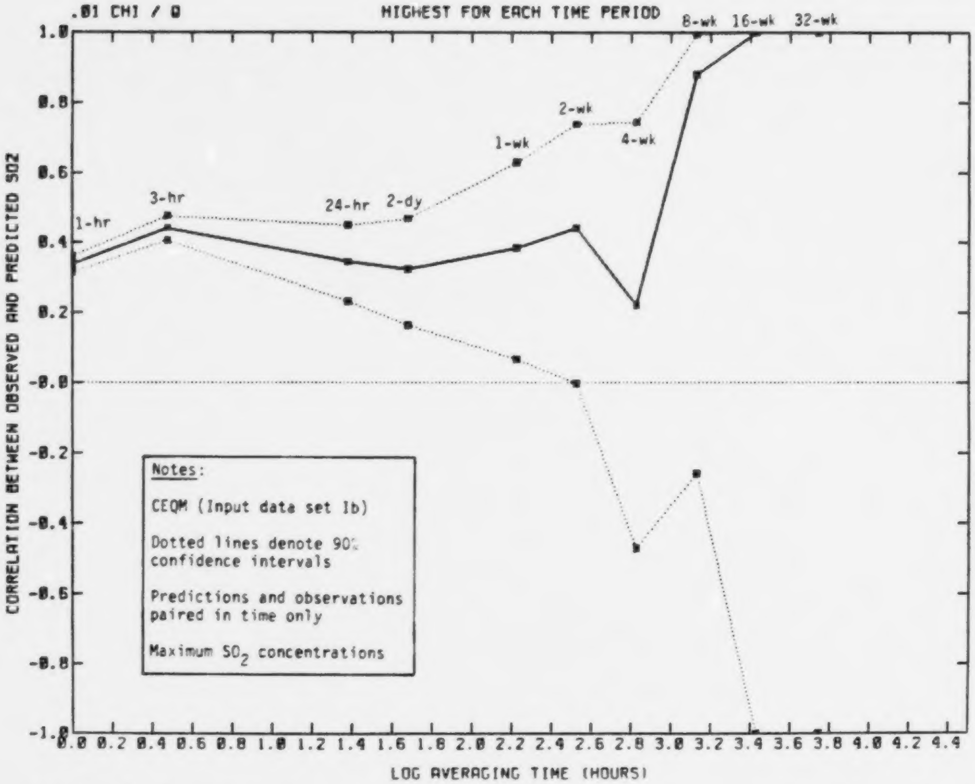


FIGURE 6-2. Dependence of time correlation on averaging time for EPRI-PMV Kincaid study (maximum SO_2 concentrations; predictions and observations paired in time only). (From Hayes et al., 1984).

tions for just the hours with the 50 highest observed concentrations are not significantly different from zero except for the PPSP and RTDM models (Table 4-1). Also, the number of predictions falling within a factor of two of the observations is less than 32 percent except for RTDM and PPSP (Table 4-1).

Table 6-2 summarizes model performance in terms of bias for time-paired predictions corresponding to the 50 highest observed concentrations. Only those data sets for which information is available are shown in this table. All models exhibit a significant underprediction bias for the time-matched residuals. This bias can be interpreted as a direct consequence of the poor correlation between observed and time-matched predicted concentrations. In other words, since observations and predictions appear to be essentially independent of one another, the highest observations will tend to occur in conjunction with predicted concentrations that are similar in magnitude to the average predicted concentration. Since this concentration is much lower than the highest observed values, an underprediction bias must exist for the hours with the highest observed concentrations. Similarily, an overprediction bias exists for the 50 hours with the highest predicted concentrations.

It has been suggested by several investigators (Ellis and Liu, 1980; Turner and Irwin, 1983; Pierce, 1988) that the poor model skill described above may be due in part to errors in the predicted direction of travel of the plume (i.e., wind direction errors). Such errors may lead to discrepancies between the highest observed and predicted concentrations for a given hour when these values are defined over a sparse network of monitoring sites since either the predicted plume may not impact any monitor during an hour when the observed plume does or vice versa. We note, however, that the use of relatively dense networks of SF_6 monitors at the EPRI PMV Kincaid site lead to only slightly better correlations between predictions and observations than those found for the more sparse SO_2 monitoring networks (Reynolds et al., 1984; Tables 6-1 and 6-15). This suggests that the poor correlations are largely the result of other factors, although the sparsity of monitoring networks no doubt plays some role.

On the basis of the above findings, we conclude that the poor model skill for time-paired observations and predictions is primarily the result of two factors:

1. Meteorological inputs to the dispersion models do not reliably represent atmospheric conditions occurring in a particular hour.

2. Model formulations are insufficient to describe the actual plume behaviors that lead to the highest observed concentrations.

We performed several analyses in an attempt to acertain the relative importance of the above factors. Results of these analyses are sumarized in the following section.

TABLE 6-2. Summary of model performance for the hours with the 50 highest observed concentrations: Observed and predicted concentrations matched in time but not location. (N/A = results not available; table compiled from results in Table 4-1).

	CRSTER/MPTER/ CEQM	TEM	RTDM	SHORTZ	PPSP
EPA Power Plant Study (Clifty Creek)		N/A	N/A	N/A	N/A
Bias[1]	Under				
Magnitude (%)[2]	79				
No. of Hours[3]	~7900				
Maryland Power Plant Siting Program		N/A	N/A	N/A	
Bias	Under				NS
Magnitude (%)	62				-22
No. of Hours	145				145
EPRI PMV					
Kincaid (SO_2)[4]			N/A	N/A	N/A
Bias	Under	Under			
Magnitude (%)	87	85			
No. of Hours	1546	1702			
Kincaid (SF_6)[4]			N/A	N/A	N/A
Bias	Under	Under			
Magnitude (%)	86	78			
No. of Hours	128	218			
Bull Run		N/A			N/A
Bias	Under		Under	Under	
Magnitude (%)	53		54	76	
No. of Hours	255		255	255	

Notes

1. Based on average residual (highest observed minus time matched highest predicted) for 50 hours with highest observed concentrations. Under = underprediction bias significant at 95 percent confidence level; NS = bias not significant at 95 percent level.
2. Ratio of average residual described in note 1 to average of 50 highest observed values.
3. Number of hours in data set from which 50 highest observations were selected.
4. Based on results for best available on-site meteorological data (input level II).

MODEL PERFORMANCE UNDER SPECIFIC DISPERSION CONDITIONS

One concern with using presently available dispersion models to perform attainment demonstrations for an hourly average ambient standard is that these models may not adequately reproduce all of the various types of dispersion phenomena that can lead to high hourly average concentrations. We performed several analyses in an attempt to identify dispersion conditions and plume behaviors that are associated with high observed concentrations and to estimate how well the models perform under such conditions.

Plume Behaviors and Dispersion Conditions Potentially Resulting in High Concentrations

A great deal of meteorological and aerometric data is needed to adequately explain the specific dispersion conditions and plume behaviors that lead to a high observed hourly average concentration. No model evaluation study carried out to date has involved the collection of enough data to allow one to unequivocally group hours with high observed concentrations into a handful of well-defined categories of dispersion conditions and plume behaviors. However, as pointed out in Section 1, we feel that enough information from previous theoretical, field, and laboratory studies is now available to make it possible to describe several general types of plume behaviors and dispersion conditions that can, and in some cases have actually been observed to, lead to high ground-level concentrations or large discrepancies between observed and predicted concentrations in the vicinity of elevated, buoyant point sources. We briefly describe these generalized plume behaviors in the following paragraphs.

The dispersive behavior of a plume is determined by several factors including:

(1) The lengthscale of the most active eddies (λ_d') relative to the width and thickness of the plume cross section (σ_y and σ_z, respectively).

(2) The intensity of turbulent mixing and the nature of the physical mechanisms producing the turbulent motions (i.e., convective or mechanical).

(3) The equilibrium height of the buoyant plume relative to the mixing height.

The size of the dominant eddy determines whether a plume gradually diffuses as it travels down wind, is torn apart into relatively large pieces, or remains relatively intact but is moved as a whole horizontally and vertically. Eddies with length scales much smaller than the cross-sectional size of the plume ($\sigma_z/2$) act to gradually diffuse the plume, while those with length scales comparable to the cross-sectional size act to tear the plume apart into relatively large pieces. Plumes may also get carried along virtually intact by big eddies (up- and downdrafts) that are much larger than the plume cross section. Plumes that are affected primarily by such large eddies appear to an observer to be moving up and down in a "looping" motion.

An important feature of the dispersion of buoyant plumes is the extent to which they are affected by the interface between unstable (or neutral) and stable air at the mixing height. Under certain ambient conditions a plume with sufficient buoyancy can rise up to or even through the mixing height, while under other conditions interaction of the plume with the mixing height is minimal. The extent to which a plume may penetrate into the air above the mixed layer is determined by the vertical velocity of the plume centerline as it encounters the top of the mixed layer. For buoyancy-dominated plumes (large SO_2 point sources do not typically have momentum-dominated plumes), the buoyancy flux of the plume (F) can be combined with the average horizontal wind speed, a nondimensional vertical velocity, and the thickness of the mixed layer to obtain a dimensionless buoyancy parameter (F_*) as described in Chapter 5. This parameter provides a quantitative measure of the potential of a plume to rise up to and penetrate the mixing height. Small values of F_* (< 0.02) imply that the plume is rising to an equilibrium height well below the mixing height, while large values ($F_* > 0.09$) suggest that the plume is penetrating complety into the stable air above the mixing height.

The size of the dominant turbulent eddies and their intensity is determined by the nature and strength of the physical forces producing them. Mechanically generated turbulence dominates where the atmosphere is in thermal equilibrium (neutral conditions), while convective turbulence occurs as a result of buoyancy forces operating in a thermally unstable atmosphere. Thermally stable stratifications act to suppress turbulent motions. The relative strengths of the buoyant and mechanical forces responsible for eddies can be estimated in a simple manner, based on routinely available meteorological data using the Pasquill-Gifford-Turner (PGT) stability classification system (Turner, 1970). PGT class A stability represents conditions in which convection dominates the turbulence structure, while D stability represents neutral conditions in which mechanically generated turbulence dominates.

A more direct measure of the relative importance of convective and mechanically generated turbulence is provided by the ratio (u/w_*) of the mean horizontal wind speed to the convective scaling velocity. Pierce (1988) assigned this ratio to categories as follows:

$u/w_* < 1.5$	Extremely unstable
$1.5 < u/w_* < 6.0$	Convective
$u/w_* > 6.0$	Not convective

Several types of plume behavior associated with high ground-level concentrations and large discrepancies between observed and predicted concentrations can be indentified (see, for example, Figure 1-1). Each behavior is typically characterized by a particular combination of PGT stability, u/w_*, and F_*, as indicated in Table 6-3.

TABLE 6-3. Description of several generalized plume behaviors potentially associated with high ground-level concentrations and/or large modeling errors.

Plume Behavior	PGT Stability Class	$\bar{u}/w_*$	F_*	λ_d'	Comments
Convective fumigation	A-B	1.5 - 3.0	<0.09	$\sigma_z/2 - z_i$	Vigorous turbulent mixing over a range of length scales brings material rapidly to the ground.
Moderate wind convection	B-C	3.0 - 6.0	0.02 - 0.07	$\sim\sigma_z/2$	Largest turbulent eddies are surpressed but rapid mixing at plume size scales is tearing the plume apart.
Bumping/looping	A-B	<1.5	0.09 - 1.0	$\sim z_i$	As in convective fumigation but with the largest eddies dominating. Sufficient buoyancy exists for the plume to reach the mixing height. May be associated with large modeling errors (see text).
Limited rise	C-D	>6.0	<0.02	$\lesssim\sigma_z/2$	High winds limit plume rise, producing high ground-level concentrations.
Plume penetration	B-D	1.5 - 6.0	>0.05	$<\sigma_z$	Sufficient buoyancy exists for the plume to partially or fully penetrate the stable layer above the mixing height. Large modeling errors may result from use of inaccurate plume penetration algorithms.

The plume behaviors identified in Table 6-3 are intended simply as a rough, and not necessarily all-inclusive, categorization of the wide range of behaviors that can be observed in nature. We have focused here on general types of behaviors that have been identified by ourselves and other investigators as leading to high hourly average concentrations or large discrepancies between observed and predicted concentrations. These behaviors are not necessarily perfectly distinct from one another, and the parameter values shown in Table 6-3 are intended only to provide a rough indication of the approximate conditions associated with each behavior.

Three of the plume behaviors identified in Table 6-3 are associated with convectively unstable conditions, one is associated primarily with neutral conditions, and a fifth may occur under a variety of stabilities. Convective conditions are especially important since they often produce high ground-level concentrations in the vicinity of elevated sources because of the rapid mixing of relatively undiluted portions of the plume to the ground near the source. Each of the plume behaviors summarized in Table 6-3 are described briefly in the following paragraphs.

Convective Fumigation

A principal feature of convective conditions is the rapid mixing of plume material to the ground (fumigation) by eddies with a wide range of length scales that simultaneously move large sections of the plume around and disperse the plume into many smaller pieces. Plumes undergoing such dispersion often appear to be "looping" up and down (see Figure 1-1a). This action results in high ground-level concentrations. Such conditions are associated with PGT stability classes A and B and light winds that are roughly of the same magnitude as the convective scaling velocity. Selecting a range of u/w_* representative of convective fumigation involves some arbitrary choices. Convective fumigation represents a moderately unstable convective condition that does not involve a predominant overturning of air on a large scale. Thus, small values of u/w_* (i.e., values less than about 1.5) are excluded from this category while larger values (greater than about 3.0) are more indicative of the moderate wind convection condition described below.

Convective fumigation differs from the other convective plume behaviors listed in Table 6-3 in that the plume buoyancy is low enough and the mixing height is high enough so that the plume is not constrained by the elevated inversion. According to Hanna, et al. (1987), a scaled buoyancy flux of greater than 0.09 indicates either plume penetration into the inversion or bumping of the plume against the inversion base (these conditions are described separately below). As a result, convective fumigation occurs only when F_* is less than approximately 0.09.

Moderate Wind Convection

If convective conditions occur in conjunction with a moderate wind, the dominant eddy length scales may be reduced to a size comparable to that of the plume cross section. The moderate wind will also reduce the plume rise, potentially to the degree that the plume centerline does not reach the top of the mixed layer. According to Willis and Deardorff (1987), this requires that the scaled buoyancy flux be substantially less than 0.14. Such conditions are consistent with PGT stability classes B - C and, according to Pierce (1988) and Weil and Brower (1984), are characterized by the maximum values of u/w_* under which convection can occur ($3.0 < u/w_* < 6.0$).

The reduced plume rise associated with moderate wind convection may result in relatively high ground-level concentrations inspite of the increased dilution brought about by the higher wind speeds.

Bumping - Looping

Under certain synoptic conditions (most commonly a strong subsidence inversion and clear skies associated with a strong high-pressure system), extremely unstable conditions (PGT stability classes A and B) may exist in a relatively shallow layer below the elevated inversion base. A low mean wind speed coupled with a large convective scaling velocity in such a mixed layer results in mixing that is dominated by large eddies with characteristic length scales on the order of the mixing height. Near the source these eddies are much larger than the plume cross section. As a result, the plume is advected vertically more rapidly than it is diffused and may appear to an observer to be looping in a vertical plane (ASME, 1973). Plume rise is enhanced by the low wind speed and, if the plume is sufficiently bouyant or else is initially entrained into an updraft, it will rise to the top of the mixed layer. Under the right combination of plume buoyancy and inversion strength, the plume may overshoot the inversion base slightly and then sink again. It has been suggested by Briggs (1985) that several such oscilliatory cycles may take place as the plume travels downwind, trapping the plume in the mixed layer and making it appear to "bump" against the inversion base. This plume behavior is illustrated in Figure 1-1c.

According to Weil and Brower (1984), bumping-looping conditions occur under low u/w_* ratios (on the order of 1.5 or less), while the corresponding PGT stability class range is A-B. The scaled buoyancy flux is greater than 0.09 under such conditions but not so large as to produce total plume penetration. Pierce (1988) suggests that F_* has an upper limit of 1.0 for bumping plumes.

So long as the plume does not lose material to the stable layers above the inversion, bumping-looping can lead to high ground-level concentrations for the same reasons they occur under convective fumigation. Bumping-looping behavior is critical to model performance because, under such conditions, models that use oversimplified algorithms for estimating the amount of plume material that remains imbedded above the inversion base or that make inaccurate estimates of plume rise and mixing height may produce highly inaccurate estimates of ground-level concentrations.

Limited Rise

While elevated, buoyant point sources do not normally produce peak ground-level concentrations under neutral conditions ($u/w_* < 6.0$, PGT stability C-D), it has been suggested by Venketram and Paine (1984) that high concentrations may occasionally result under such conditions if the wind speed is sufficiently high to severly reduce plume rise and increase the thickness of the mechanically mixed layer. If the bouyancy flux is low, only a moderate wind speed may be needed to reduce the plume rise to the point that high ground-level concentrations are observed. The scaled buoyancy flux under these conditions will be small, and plume dispersion will be carried out by eddies generated by wind shear. The highest ground-level concentrations will occur when the typical eddy size is on the order of the size of the plume cross section, resulting in rapid mixing of the plume to the ground. This plume behavior is illustrated in Figure 1-1b.

Plume Penetration

Penetration of the plume into the stable air above the base of the inversion can occur under a variety of atmospheric conditions. The distinguishing feature of this phenomenon is that the plume rise is sufficient to either partially or fully bring the plume above the mixing height ($F_* > 0.05$). Large prediction errors can result under such conditions if a model with an oversimplified penetration algorithm is employed. Errors in the proper specification of plume rise and mixing height can also contribute to modeling errors under these conditions.

Inversion Break-up and Shoreline Fumigation

Inversion break-up and shoreline fumigation have been described by Turner (1970), Budney (1977), and Lyons and Cole (1973) among others. These phenomena can lead to high ground-level concentrations at certain sites. They involve the release of a plume into an elevated stable layer within which dispersion is minimized, thus allowing high centerline concentrations to persist for considerable distances downwind. At some point, these high concentrations are mixed rapidly to the ground either through a morning inversion break-up process or because a plume initially released into a stable (e.g., marine or nighttime rural) environment has crossed over into a region of instability (e.g., a plume released into a stable marine air mass that is gradually eroded from below as it passes over land or a plume released into a stable, nighttime rural air mass that travels over an urban area and is made unstable by the urban heating and increased surface roughness). This plume behavior is illustrated in Figure 1-1d.

Applicability of Model Formulations to Important Plume Behaviors

In Chapter 3 we reviewed model formulations and commented on their applicability to the various dispersion phenomena associated with high ground-level concentrations. We found that no presently available point source dispersion model contains all of the algorithms needed to make accurate predictions under all of the plume behaviors described above. There are, however, sufficient differences between the models included in our review to suggest that some may be better equipped to deal with these phenomena than others. For example, plume behaviors such as bumping-looping and partial penetration in which the plume centerline is near the mixing height can prove to be difficult for models to reproduce accurately unless some sort of partial penetration algorithm is included in the model. Neither the CRSTER/MPTER/CEQM or SHORTZ models have such an algorithm. These models assume a zero ground-level impact if the plume centerline is predicted to be above the mixing height and a perfect reflection of plume material off of the inversion base otherwise. This oversimplified approach can lead to large prediction errors, as demonstrated in a later section.*

Accurate estimates of plume height during convectively unstable conditions are needed to produce accurate predictions of ground-level impacts. The plume breakup/touchdown algorithm in the PPSP model represents an improvement over the standard Briggs formulas used in the other models that is designed to meet this need.

Models that rely solely on the PGT stability class to estimate the rate of plume dispersion may have difficulty differentiating between the effects of various convective and mechanically driven turbulence phenomena and may use inaccurate dispersion coefficients for some hours since the PGT classes are based solely on wind speed, cloud cover, and solar zenith angle. Some of the models we reviewed (RTDM, SHORTZ and PPSP) have the capability of using more direct measures of turbulence intensity (such as u/w_*) in calculating dispersion coefficients. If accurate estimates of turbulence intensities are available, these models should produce more realistic dispersion coefficients and concentration predictions.

---

* Although TEM and RTDM have partial penetration algorithms, these models were exercised in their respective model evaluation studies with the algorithm inactivated. When operated in this mode, these models assume that the plume centerline never rises above the mixing height and that there is no "reflection" of plume material at the mixing height.

Dispersion Conditions Associated with Highest Observed Concentrations

As discussed above, on the basis of theoretical and laboratory studies we have identified several dispersion conditions that have the potential to produce high observed ground-level concentrations in the vicinity of a buoyant, elevated point source. Some of these conditions, such as inversion break-up fumigation, occur only occasionally at isolated locations and last only briefly, making them difficult to observe in the field. Other conditions, such as those which result in a bumping-looping plume, are difficult to distinguish from conditions that share many of the same characteristics (PGT stability classes A-B) but may be associated with quite different model performance results (e.g., convective fumigation). Therefore, it is often difficult to identify which particular dispersion condition or plume behavior is associated with a particular high observed concentration. Nevertheless, our review of previous model performance evaluation studies and analysis of meteorological conditions associated with the highest observed concentrations for a few selected studies (Tables 4-4, 5-1 and 5-2) does allow us to make some general observations about the dispersion conditions associated with peak ground-level impacts.

As expected, virtually all of the highest concentrations observed during the studies we reviewed occurred under unstable conditions. Two exceptions to this general rule were found for the EPRI PMV Kincaid site (Tables 4-4 and 4-5) where two hours of high nighttime concentrations were observed in conjunction with high winds at stack height. No instances of high impacts due to limited plume rise under daytime neutral conditions could be identified, although one of the 30 hours of high concentrations around the Paradise plant analyzed by Pierce (1988) appears to fit into this category. It was also not possible to discern from the available data any impacts associated with morning inversion break-up fumigation, although anecdotal evidence of such events occurring at the EPRI PMV Kincaid site is available. From this it appears that, while limited plume rise and fumigation situations may indeed produce occasionally high concentrations, the vast majority of high concentrations (including the very highest observed in each of the studies we examined) are associated with convectively unstable conditions.

Pierce (1988) noted that roughly two thirds of the 30 highest concentrations observed around the Paradise plant were associated with unstable conditions and ratios of plume height to mixing height ranging between 0.8 and 1.0.* Pierce points out that these heights correspond to the "entrainment layer" defined by Gryning and co-workers (1987) and suggests that the plume is bumping against (but not penetrating) the inversion base under these conditions. This plume behavior is synonymous with the bumping-looping behavior identified in Table 6-3. Pierce also analyzed the distribution of F_* for the 30 highest hours and found that only half of the values fell within the 0.09 - 1.0 range associated with bumping plumes; the remaining half had smaller values. These proportions were roughly identical to those found when all 8754 hours in the data set were examined. Our own analysis of the 50 hours with the highest observed concentrations at four different sites found that the median values of F_* fall in the range 0.10 to 0.20. Taken together, these results suggest that the highest observed concentrations frequently arise when a buoyant plume is trapped under a strong inversion and convective activity below the inversion mixes relatively undiluted material to the ground.

---

* Plume heights estimated by methods proposed by Briggs (1984); mixing heights obtained from inspection of twice-daily soundings.

The remaining one-half to one-third of the 30 high concentration events analyzed by Pierce (1988) that do not appear to be associated with bumping-looping but nevertheless occurred under convective conditions are cases of either convective fumigation or moderate wind convection. As indicated in Table 6-3, the principal difference between these two phenomenon is the length scale of the dominant eddies as determined at least in part by the mean horizontal wind speed and the convective scaling velocity (u/w_*). Pierce's results for u/w_* do not distinguish between values greater than or less than 3.0. However, based on the separate distributions of u and w_* reported by him, one can see that nearly all of the values of u/w_* during the 30 highest hours must be less than 3.0, suggesting that the hours in question fall in the convective fumigation category. Even discounting the estimates of w_* obtained by Pierce, the fact that 23 of the 30 highest hours had wind speeds of less than 4 m/s suggests that convective fumigation, not moderate wind convection, was taking place during these hours. Thus, high observed concentrations can also occur when plume heights are well below the mixing height and large-scale eddies rapidly mix plume material to the ground.

Dispersion Conditions Associated with Over- and Underpredictions

From our review of the results of previous model evaluation studies (presented in Chapter 2) it seems clear that, as a general rule, model predictions of the N highest concentrations are biased toward overprediction under unstable conditions while many models exhibit an underprediction bias under neutral conditions. Residuals appear to be mixed under stable conditions but there is generally little data available for such conditions. Many of the same studies described in Chapter 2 were analyzed by Hayes and co-workers (1984) who showed that the mean ratio of predicted to observed N highest concentrations unmatched in time or location usually decreases with increasing stability, often going from a value greater than 1.0 under unstable conditions to a value less than 1.0 under neutral, and in many cases stable, conditions. These results are summarized in Figures 6-3 to 6-5.

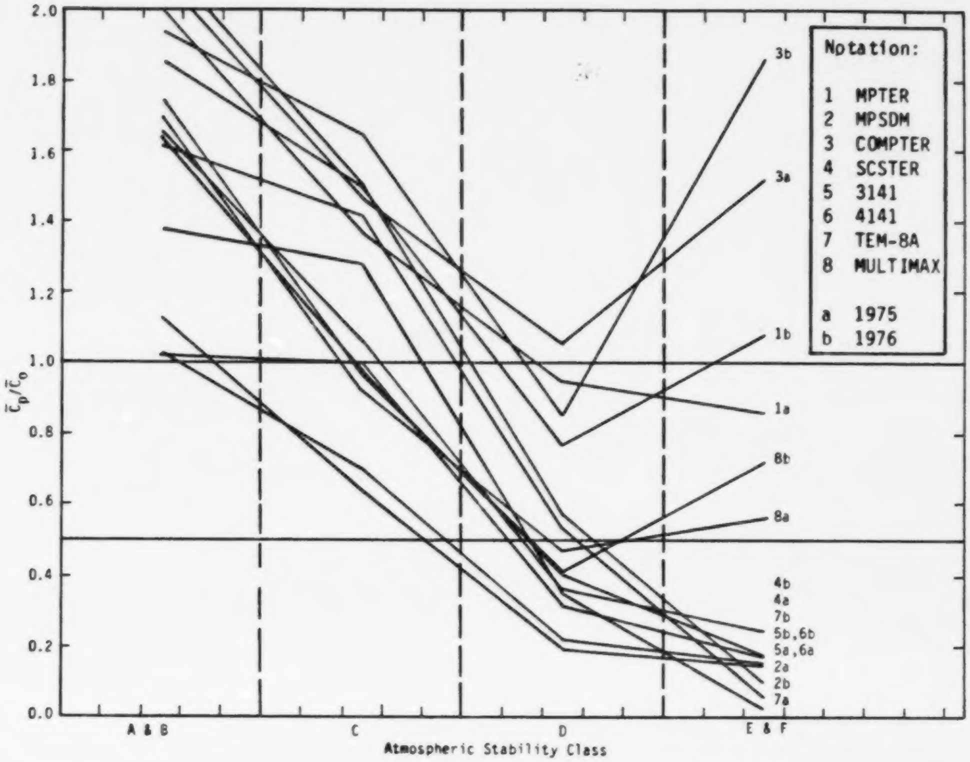
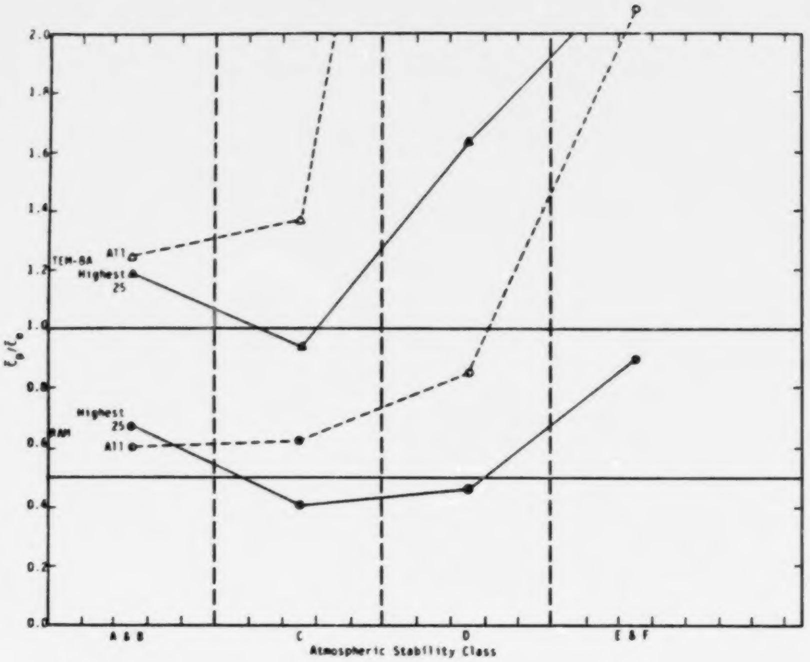
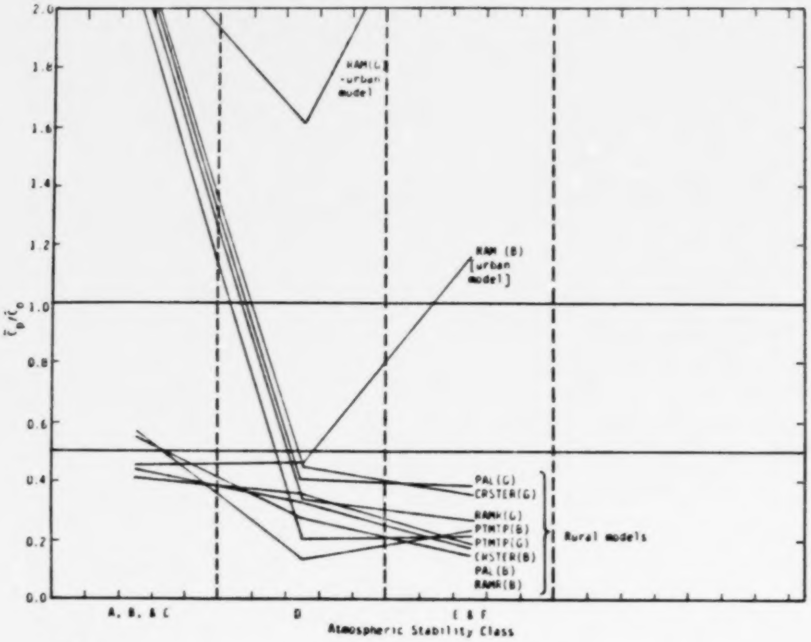


FIGURE 6-3. Dependence of model bias on atmospheric stability class in the EPA Rural Study (Clifty Creek 1975 data). Ratios of means of highest 25 1-hour SO_2 averages. (From Hayes et al., 1984).



Notes:

FIGURE 6-4. Dependence of model bias on atmospheric stability class in the EPA Urban Study (St. Louis 1976 data). Ratios of means of predicted and observed 1-hour SO_2 averages (for the highest 25 unpaired values; and all values, paired in space and time). (From Hayes et al., 1984).



Notes:

FIGURE 6-5. Dependence of model bias on atmospheric stability class in the API Tracer Study. Ratios of means of highest 10 predicted and observed 1-hour averages. The data are from two elevated release, buoyant plume tracer tests (both urban): G = Goodyear, B = Benecia. (Source: Hayes et al., 1984).

Our analysis of the 50 highest time-paired observed and predicted concentrations at four sites (Table 5-4) revealed a positive correlation of hourly residuals with stability class (i.e., predicted values decreased with respect to observations as stability increased) except at the EPRI PMV Kincaid site and for the PPSP model from the Maryland Power Plant Siting study (where almost no A stability hours were observed). Scatter plots of residuals versus stability class (Figures 5-1 through 5-5) reveal that, with the exceptions noted above, during the hours with the 50 highest observations the largest overpredictions tended to occur under unstable conditions whereas underpredictions were the general rule during neutral and stable conditions (when observed). This association between residuals and stability class is further confirmed by the results of our cluster analysis, in which a standard clustering algorithm was used to sort hours into groups with similar residuals and meteorological conditions. Clusters consisting mostly of hours with overpredictions were identified with very unstable conditions, while underprediction clusters included hours with greater stabilities (see, for example, Table 5-5). These results are consistent with our finding of an underprediction bias for the hours with the 50 highest observations. The average residual is an underprediction and the few overpredictions that do occur must be associated with meteorological conditions (A and B stabilities) that produce high predicted concentrations. Nearly neutral and stable conditions (PGT classes C through F) generally result in lower predicted concentrations; thus, such hours, which occur more frequently in many data sets, are the major contributors to the underprediction bias.

Although some model evaluation studies we reviewed included an examination of the relationship of residuals to other meteorological variables, such as wind speed and mixing heights, significant associations were generally not found. Our own comparison of meteorological conditions associated with the highest overpredictions and those associated with the largest underpredictions (Tables 4-5 and 4-6) did find that mixing heights tended to be lower during the underprediction hours. This was further confirmed by our cluster analysis (Tables 5-5 through 5-8) for models other than PPSP and RTDM. Clusters consisting of underpredictions are generally associated with lower-than-average mixing heights. Scatter plots of the hours with the highest over- and underpredictions, such as the one in Figure 4-3, show that almost all of the underpredictions result when a zero impact is predicted even though a relatively high concentration is observed during the same hour. These findings suggest that the largest underpredictions for models other than PPSP and RTDM are primarily the result of incorrect plume rise treatment during strong inversion conditions indicative of bumping/looping behavior. That is to say, the models are predicting that the plume is rising through the inversion and not impacting the ground when in fact the plume's rise is being suppressed by the inversion and significant amounts of material are being mixed to the ground.

Our analysis of meteorological variables associated with the 50 highest observed concentrations (Table 5-4 and Figures 5-1 through 5-5) did not reveal any consistent significant associations of residuals with variables other than stability class and mixing height. In particular, we were unable through this limited analysis to identify any relationship of residuals with either u/w_* or F_*. Our cluster analysis also failed to find any outstanding associations of residuals with variables other than stabilitiy class and mixing height. However, we note that our analysis was confined to the hours with the 50 highest observed concentrations. It is possible that significant relationships could be found if a larger data set, e.g., several hundred hours, were to be used.

For those model evaluation studies for which observations are from a very sparse network of monitors (e.g., Clifty Creek), underpredictions of the type illustrated in Figure 4-3 may also result from errors in the wind direction used by the model to predict the direction of plume travel. Large underpredictions can be generated for hours during which the wind is actually blowing material toward a monitor (thus resulting in a non-zero observation) but the wind direction used by the model is such that no monitor is predicted to be impacted by the plume. However, underpredictions consisting of zero predictions matched with significant observations were found to occur even at locations with dense monitoring networks (i.e., Kincaid and Bull Run), thus suggesting that wind direction error is not necessarily the leading cause of such underpredictions.

Our analysis of the largest overpredictions at selected sites (Table 4-6) shows that overpredictions generally occur under unstable light wind conditions when the mixing height is relatively high, i.e., convective fumigation conditions. Under such conditions the predicted plume height is below the inversion base and A or B stability dispersion coefficients are assumed. As a result, substantial ground-level impacts are predicted. The fact that observed impacts are lower than the predicted values under these conditions suggests that the A and B stability dispersion coefficients and reflection algorithms used by most of the models may be inaccurate. It has been suggested by Willis and Deardorf (1978) that the vertical distribution of a buoyant plume under convective conditions is non-Gaussian and asymetric with lower concentrations at a given distance below the plume centerline than above. This asymetry is believed to be the result of the modifying effects of the plume's buoyancy on its dispersion (Lamb, 1984) and may explain the overpredictions. Another contributing factor may be that some of the plume material is being trapped in the elevated inversion. The perfect reflection algorithm used by most models does not recognize this possibility, thus producing ground-level concentrations that are too high.

SUMMARY AND CONCLUSIONS

On the basis of the model performance analyses described above, it is possible to make some general statements about the ability of presently available dispersion models to accurately predict peak hourly average concentrations in the vicinity of elevated, buoyant point sources. We summarize our findings in the following sections. This is followed by a brief discussion of the implications of these findings for regulatory applications of presently available dispersion models in connection with a 1-hour SO_2 standard.

Model Performance: Peak Concentrations Unpaired In Time or Location

In general, the CRSTER, MPTER, and CEQM models tend to overpredict the average of the N highest (N = 25 in most cases) observed concentrations when observations and predictions are unpaired in time or location. The magnitude of the overprediction was found to be no more than about 50 percent in most cases. TEM also overpredicted slightly at most sites. PPSP and MPSDM produced much larger overpredictions at all sites. Overpredictions of the highest and second-highest observed concentrations were also common, although results for these extreme values are not as consistent across studies as for the average of the N highest, thus reducing the reliability of any generalizations about model performance for these extreme values.

Comparison of Hourly Average Performance With 3- and 24-Hour Average Performance

The average of the N highest predicted concentrations was found to decrease relative to the observed value as averaging time was increased from 1 to 24 hours. Models that overpredict hourly averages tend to underpredict or only slightly overpredict 24-hour averages. Correlations between maximum observed and predicted concentrations matched in time but not location do not improve significantly with longer averaging periods.

Model Performance: Peak Concentrations Matched in Time Only

Only the Maryland/PPSP and Bull Run/RTDM model evaluations showed appreciable model precision as measured by the correlation between maximum observed and predicted concentrations matched in time. The CRSTER, MPTER, CEQM or TEM models produced very low correlations in most cases. Correlations obtained from dense tracer monitoring networks were not significantly better than those obtained from relatively sparse SO_2 monitoring networks, showing that wind direction errors are not the major cause of the poor model precision for time-matched pairs of observations and predictions. Consistent with this lack of predictive skill, models were found to underestimate the average of the 50 highest observed concentrations when predictions and observations were matched in time. This was found to be true even of the PPSP and RTDM models.

Model Performance for Specific Dispersion Conditions

Based on our review of theoretical and laboratory experiments, we identified several types of plume behaviors that have been postulated as significant to the prediction of peak hourly average concentrations. These behaviors are associated with specific dispersion conditions that can be characterized by a group of scaling parameters (u/w_*, F_*, λ'_d) and the PGT stability class. Three convective behaviors were identified ("convective fumigation", "moderate wind convection" and "bumping-looping") along with a high wind, neutral condition ("limited rise"). A fifth behavior ("plume penetration") can occur under a variety of atmospheric conditions. With the exception of plume penetration, all of these behaviors have the potential of producing high ground-level concentrations.

We briefly reviewed the ability of model formulations to describe the significant plume behaviors and dispersion conditions we identified. Two major areas in which the CRSTER and similar models were found to be lacking were:

> An inability to model the reduction in plume rise that can occur as the result of the presence of a strong elevated inversion.
>
> Reliance on the PGT stability classification system to distinguish between different convective dispersion conditions.

Formulations of the PPSP and RTDM models were specifically designed to address these concerns.

We reviewed the results of previous model evaluations and performed an independent analysis of data obtained for all studies for which it was available. On the basis of our review and anslysis, we made the following observations:

> The highest observed concentrations almost always occur under unstable conditions. A few isolated instances of high concentrations occurring under conditions of high wind speed and limited plume rise were noted but these concentrations were never among the very highest observed (i.e., in the top 10).
>
> The anlysis by Pierce (1988), although confined to a single source, shows that bumping-looping and convective fumigation conditions play a significant role in high observed concentrations.
>
> Models tend to overpredict the maximum concentration under unstable conditions unless the predicted plume height is greater than the mixing height, in which case the CRSTER and similar models strongly underpredict (i.e., predict a zero impact). When the N highest observed and predicted concentrations unpaired in time or location are stratified by stability, a strong overprediction bias is found for unstable conditions (since the inappropriate zero predictions are not included in such a data set). This suggests that the bumping-looping behavior may be associated with underpredictions if the plume rise algorithm used in the model does not properly account for the presence of the elevated inversion while the convective fumigation and moderate wind convection conditions are associated with overpredictions.

Stratification of the unpaired observed and predicted concentrations by stability class also reveals an underprediction bias for the N highest concentrations under neutral conditions, in contrast to the overprediction bias under unstable conditions described above. The data for stable conditions are sparse and results were mixed.

Insufficient data is available to determine how well the models are able to handle limited plume rise or inversion break-up fumigation conditions. Such conditions were observed rarely, if at all, in the data sets we examined.

Implications For Model Applications In Connection With a One-Hour Sulfur Dioxide Standard

If an hourly average NAAQS for SO_2 were to be promulgated, would currently available dispersion models be suitable for use in demonstrating attainment of the standard in the vicinity of large, elevated, buoyant point sources such as utility power plants? The results summarized above suggest that using these models in connection with an hourly average standard does not present the regulatory community with any significant additional uncertainties that are not already being dealt with in connection with the current 3- and 24-hour NAAQS. Predictions of peak hourly average concentrations unmatched in time or location are slightly more conservative on average than corresponding predictions of 3- and 24-hour averages, and model precision for observations and predictions matched in time is no worse for hourly averages than it is for longer averaging times.

Although certain conditions not specifically treated by the models, such as inversion break-up and seabreeze fumigation, are known to occur and to have the potential for producing high ground-level impacts from elevated sources, we found that these conditions have been only rarely observed in model evaluation studies and that they do not seem to contribute to the bulk of the 50 highest concentrations. In fact, nearly all of the highest concentrations observed in the evaluation studies we examined occurred under convective conditions, which are treated by most of the models as PGT stability class A and B cases. Peak predictions under such conditions match reasonably well with the highest observed concentrations, although large prediction errors can occur in any given hour. Thus, special dispersion situations that might be of concern in modeling peak hourly average concentrations do not appear to occur frequently enough around typical power plant sources to invalidate the use of currently available dispersion models. In reaching this conclusion, we have explicitly excluded consideration of situations involving impacts on elevated terrain or building downwash phenomena.

RECOMMENDATIONS

Our review of the results of model evaluation studies was restricted to some fairly simple analyses. In the following section we present some recommendations for extending these analyses to provide a clearer picture of the reasons for the observed over- and underpredictions. This is followed by a brief set of recommendations for improvements to model formulations that our findings suggest would result in improved model performance.

Additional Analyses

A principal limitation of our analysis of the hours with the 50 highest observed concentrations and the 50 highest residuals is that the distributions of meteorological variables for all hours in the evaluation data sets were not available for comparison with the corresponding distributions for the 50 highest hours. Our analysis should be extended to include such a comparison, as was done, for example, by Pierce (1988) using the Paradise plant data set. In this way, the statistical significance of meteorological differences between hours with overpredictions, underpredictions, and high observed concentrations can be asessed. The use of additional techniques for organizing the multivariate data set consisting of observations, predictions, and meteorological conditions should be also be explored. In particular, we feel that a Classification and Regression Tree (CART) analysis (Breiman, et al., 1984), in which the residuals are treated as the dependent variable and the meteorological variables as the independent (predictor) variables, would provide useful insight into the key meteorological conditions associated with various types of residuals. A CART analysis with the observed concentration as the dependent variable would also be of interest.

One difficulty encountered in our analysis was the presence of outliers in computed scaling parameters, such as F_* and u/w_*. These outliers should be examined to determine their origin and validity. This would lead to an improved data set from which the conditions associated with over- and underpredictions could be better inferred.

Such a large amount of meteorological, aerometric, and remote-sensing data was collected in connection with the EPRI PMV and Maryland PPSP model evaluation studies that it is not feasible to examine all of it for a large number of hours. However, a case-by-case analysis of a few carefully selected hours with large residuals and large observed concentrations would provide much additional information on the reasons why the models behave in a particular way. We recommend that such an analysis be carried out, with particular emphasis on examination of plume structure as observed by LIDAR and COSPEC instruments.

We have noted that poor time-matched model performance may be due at least in part to the fact that model predictions represent a mean value over an ensemble of similar meteorological conditions while the observed values represent a single realization of an event from the ensemble. It may be possible to estimate the size of the contribution of this effect to the observed lack of model skill by comparing observed and predicted concentrations unmatched in time that belong to the same ensemble. Ensembles could be defined in terms of simple model inputs (e.g., all hours of A stability, very light winds, and low mixing heights).

Last, but certainly not least, we recommend that additional data analyses and possibly field studies be carried out to identify situations in which inversion break-up and seabreeze fumigation as well as downwash phenomena occur and to characterize model performance under these conditions.

Recommendations for Improvements in Model Formulations

As pointed out above, one of the principal causes of poor model performance is the apparently unrealistic treatment of plume behavior in the vicinity of the mixing height. In particular, as pointed out by Turner (1985), the overly simplified treatment of the vertical profiles of mean wind and stability parameters used in most models can lead to inaccurate estimates of plume height, especially in the vicinity of elevated inversions. This in turn can result in inaccurate estimates of ground-level concentrations. For example, the CRSTER/MPTER/CEQM and SHORTZ models assume a zero ground-level impact whenever the predicted plume height is greater than the mixing height. Given the inevitable errors present in estimates of both of these quantities and the fact that the plume rise is calculated without regard to the limitation to rise imposed by the elevated inversion, zero ground-level concentrations are often predicted when in actuality substantial impacts are occurring. That such situations often dominate cases of underprediction is made strikingly clear in Figure 4-3. Substantial reduction of the number of cases of underprediction could be achieved through the use of more realistic treatments of the interaction of plume and inversions.

Another area in which model improvements are needed is in the simulation of plume dispersion within the mixed layer during convective conditions. The commonly used PGT stability classes and dispersion parameters do not always accurately describe the degree of lateral and vertical dispersion as well as formulations based on directly observed stability parameters, such as u/w_*, or the enhanced lateral dispersion of a plume during bumping-looping as discussed by Pierce (1988).

It is unclear from our review what role errors in specifying wind direction play in model performance. Pierce (1988) and others (Ellis and Liu, 1980; Turner and Irwin, 1983) have suggested that misspecifications of wind direction often lead to large errors in model predictions even when predictions are paired only in time (and not location) with observations for sparse (low resolution) monitoring networks. While wind direction may indeed play a role in time-matched model performance, we have noted similar sorts of modeling errors at sites using a low-resolution network (Clifty Creek) and those with a high-resolution network (Bull Run tracer monitors). This suggests to us that improvements in models designed to provide a more realistic treatment of wind direction are not as important as improvements in the aforementioned areas.

References

Anderberg, M. R. 1973. Cluster Analysis for Applications. Academic Press.

ASME. 1973. Recommended Guide for the Prediction of the Dispersion of Airborne Effluents, second edition. The American Society of Mechanical Engineers, New York.

Bjorklund, J. R., and J. F. Bowers. 1982. User's Instructions for the SHORTZ and LONGZ Computer Programs, Vol. 1. U.S. Environmental Protection Agency (EPA-903/9-82-004a)

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and Regression Trees. Wadsworth, Belmont, California.

Briggs, G. A. 1970. "Some Recent Analyses of Plume Rise Observations." International Air Pollution Conference, Washington, D.C.

Briggs, G. A. 1972. Discussion of "Chimney plumes in neutral and stable surroundings." Atmos. Environ., 6:507-510.

Briggs, G. A. 1973. "Diffusion Estimation for Small Emissions." Atmospheric Turbulence and Diffusion Laboratory, National Oceanic and Atmospheric Administration, Oak Ridge, Tennessee.

Briggs, G. A. 1975. "Plume Rise Predictions." In Lectures on Air Pollution and Environmental Impact Analysis, D. A. Haugen, ed., American Meteorological Society, Boston, Massachusetts.

Briggs, G. A. 1984. "Plume Rise and Buoyancy Effects." In Atmospheric Science and Power Production. Atmospheric Turbulence and Diffusion Laboratory. National Oceanic and Atmospheric Administration, Oak Ridge, Tennessee.

Briggs, G. A. 1985. Analytical parameterizations of diffusion: the convective boundary layer. J. Climate Appl. Meteorol., 24(11):1167ff.

Budney, L. J. 1977. Guidelines for Air Quality Maintenance Planning and Analysis Volume 10 (Revised): Procedures for Evaluating Air Quality Impact of New Stationary Sources. U.S. Environmental Protection Agency (EPA-450/4-77-001); OAQPS No. 1.2-029 R).

Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley. 1971. Flux-profile relationships in the atmospheric surface layer. J. Atm. Sci., 28:181-189.

Busse, A. D., and J. R. Zimmerman. 1973. User's Guide for the Climatological Dispersion Model. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-RA-73-024; NTIS PB 227346/AS).

Cox, B. 1983. "EPA Model Evaluation Program." Third Conference on Air Quality Modeling, Washington, D.C.

Cox, W. M., and G. K. Moss. 1985. Evaluation of Rural Air Quality Simulation Models. Addendum A: Muskingum River Data Base. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/4-83-003a).

Cox, W. M., H. W. Rorex, and G. K. Moss. 1986. Evaluation of Rural Air Quality Simulation Models. Addendum C: Kincaid SO2 Data Base. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/4-83-003c).

Cox, W. M., H. W. Rorex, and G. K. Moss. 1987. Evaluation of Rural Air Quality Simulation Models. Addendum D: Paradise SO2 Data Base. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/4-83-003d).

Ellis, H. M., and P. C. Liu. 1980. Discussion of "An air quality model performance assessment package." Atmos. Environ., 14:1113.

Ellis, H. M., and P. C. Liu. 1981. Review of the performance of the RAM model in predicting highest measured concentrations. J. Air Pollut. Control Assoc., 31:(2):148-152.

EPA. 1977. User's Manual for Single Source (CRSTER) Model. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/2-77-013; NTIS PB-271360).

EPA. 1986. Guideline on Air Quality Models (Revised). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/2-78-027R).

ERT. 1982. "User's Guide to the Rough Terrain Diffusion Model, Rev. 3.00." Environmental Research and Technology, Inc., Concord, Massachusetts (M2209-585).

Fox, D. G. 1981. Judging air quality model performance. Bull. Am. Meteorol. Soc., 62:599-609.

Gifford, F. A. 1961. Uses of routine meteorological observations for estimating atmospheric dispersion. Nuclear Safety 2:47-51.

Gifford, F. A. 1972. Atmospheric transport and dispersion over cities. Nuclear Safety, 13:391-402.

Gifford, F. A. 1975. "Atmospheric Dispersion Models for Environmental Pollution Applications." In Lectures on Air Pollution and Environmental Impact Analysis, American Meteorological Association, Boston.

Gryning, S. E., A.A.M. Holtslag, J. S. Irwin, and B. Sivertsen. 1987. Applied dispersion modeling based on meteorological scaling parameters. Atmos. Environ., 21:79-89.

Guldberg, P. H., and C. W. Kern. 1978. A comparative validation of the RAM and PTMTP models for short-term SO_2 concentrations in two urban areas. J. Air Pollut. Control Assoc., 28(9):907-910.

Hanna, S. R., G. A. Briggs, J, Deardorff, B. A. Egan, F. A. Gifford, and F. Pasquill. 1977. AMS workshop on stability classification schemes and sigma curves--summary of recommendations. Bull. Am. Met. Soc., 58:1305-1309.

Hanna, S. R., J. C. Weil, and R. J. Paine. 1986. "Plume Model Development and Evaluation--Hybrid Approach." Environmental Research and Technology, Inc., Concord, Massachusetts (DO34-500).

Hanna, S. R., J. C. Weil, and R. J. Paine. 1987. "Testing of A New Plume Dispersion Model With the Kincaid and Bull Run Data." Air Pollution Control Association annual meeting, New York (paper 87-73.4)

Hayes, S. R. 1979. "Performance Measures and Standards for Air Quality Simulation Models." Systems Applications, Inc., San Rafael, California (EF78-93R2).

Hayes, S. R., and G. E. Moore. 1986. Air quality model performance: A comparative analysis of 15 model evaluation studies. Atmos. Environ., 20(10):1897-1911.

Hayes, S. R., G. E. Moore, and J. E. Langstaff. 1984. "Air Quality Model Performance: A Comparative Analysis of Model Evaluation Studies and their Implications for Federal PSD Policy." Systems Applications, Inc., San Rafael, California (SYSAPP-84/210).

Hellums, L. 1982. Rigorous Mathematical Analysis for Gaussian Plume-Representation of Tracer Field Data. Phillips Petroleum Company.

Hodanbosi, R. F., and R. Brodkey. 1984. Evaluation of RAM model for Cleveland, Ohio, Part II. J. Air Pollut. Control Assoc., 34(4):371-374.

Hodanbosi, R. F., and L. K. Peters. 1981. Evaluation of RAM model for Cleveland, Ohio. J. Air Pollut. Control Assoc., 31(3):253-255.

Holtslag, A.A.M., and A. P. van Ulden. 1983: A simple scheme for daytime estimates of the surface fluxes from routine weather data. J. Climate Appl. Meteorol., 22:517-529.

Irwin, J. S. 1979. "Estimating Plume Dispersion--A Recommended Generalized Scheme." American Meteorological Society, Boston. Fourth Symposium on Turbulence, Diffusion and Air Pollution.

Irwin, J., and M. Smith. 1984. Potentially useful additions to the rural model performance evaluation. Bull. Amer. Meteorol. Soc., 65(6):559-568.

Kummler, R. H., B. Cho, G. Roginski, R. Sinha, and A. Greenberg. 1979. A comparative validation of the RAM and modified SAI models for short term SO_2 concentrations in Detroit. J. Air Pollut. Control Assoc., 29(7):720-723.

Lamb, R. G. 1984. "Diffusion in the Convective Boundary Layer." In Atmospheric Turbulence and Air Pollution Modeling, F.T.M. Nieuwstadt and H. van Dop, eds., D. Reidel, Dordrecht, The Netherlands.

Londergan, R. J., J. J. Mangano, N. E. Bowne, D. R. Murray, and H. Borenstein. 1980. An Evaluation of Short-Term Air Quality Models Using Tracer Study Data. Volume I. American Petroleum Institute, Washington, D.C. (No. 4333).

Londergan, R., D. Minott, D. Wackter, and R. Fizz. 1983. Evaluation of Urban Air Quality Simulation Models. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/4-83-020).

Londergan, R. J., D. Minott, D. Wackter, T. Kincaid, and D. Bohitata. 1982. Evaluation of Rural Air Quality Simulation Models. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/4-83/003).

Londergan, R. J., D. R. Murray, N. E. Bowne, and H. Borenstein. 1980. "A Comparison of Predictions from Standard Short-Term Air Quality Models with Observed Tracer Dispersion." Second Joint Conference on Applications of Air Pollution Meteorology and Second Conference on Industrial Meteorology. American Meteorological Society.

Lott, R. A. 1984. Performance of a time-dependent dispersion model in elevated terrain. Atmos. Environ., 18:2367-2376.

Lyons, W. A., and H. S. Cole. 1973. Fumigation and plume trapping on the shores of Lake Michigan during stable onshore flow. J. Appl. Meteorol., 12:494-510.

McTaggart, M. 1981. "Comparative Study of ISC, MPTER, and BLP Dispersion Models." APCA Specialty Conference on Dispersion Modeling for Complex Sources, St. Louis, Missouri.

Mills, M. T., R. Caiazza, D. D. Hergert, and D. A. Lynn. 1979. "Evaluation of Point Source Dispersion Models." Teknekron Research, Inc., Waltham, Massachusetts.

Mills, M., and F. A. Record. 1975. Comprehensive Analysis of Time-Concentration Relationships and the Validation of a Single-Source Dispersion Model. U.S. Environmental Protection Agency, Washington, D.C. (EPA-450/3-75-083).

Mills, M., and R. Stern. 1975. Model Validation and Time-Concentration Analysis of Three Power Plants. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-450/3-76-002).

Paine, R. J., and S. R. Hanna. 1986. "User's guide to the Hybrid Plume Dispersion Model (HPDM) and Its Associated Processors (Version 2.1). Environmental Research and Technology, Concord, Massachusetts (PD-039-585).

Pasquill, F. 1976. Atmospheric Dispersion Parameters in Gaussian Plume Modeling. Part II. Possible Requirements for Change in the Turner Workbook Values. U.S. Environmental Protection Agency, Research Triangle Park, N.C. (EPA-600/4-76-030b).

Pierce, T. E., and D. B. Turner. 1980. User's Guide for MPTER. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina (EPA-600/8-80-016).

Pierce, T. E. 1988. An Air Pollution Climatology of An Isolated Point Source Using Convective Scaling Parameters. Atmos. Environ. (in press).

Rao, S. T., G. Sistla, V. Pagnotti, W. B. Petersen, J. S. Irwin, and D. B. Turner. 1985. Evaluation of the performance of RAM with the regional air pollution study data base. Atmos. Environ., 19(2):229-245.

Reynolds, S. D., T. C. Myers, J. E. Langstaff, M.-K. Liu, G. E. Moore, and R. E. Morris. 1985. Operational Validation of Gaussian and First-Order Closure Plume Models at a Moderately Complex Terrain Site. Elecric Power Research Institute, Palo Alto, California (EPRI EA-3759).

Reynolds, S. D., C. Seigneur, T. E. Stoeckenius, G. E. Moore, and R. G. Johnson. 1984. Operational Validation of Gaussian Plume Models at a Plains Site. Electric Power Research Institute, Palo Alto, California (EPRI EA-3076).

Ruff, R. E. 1980. Evaluation of the Real-Time Air-Quality Model Using the RAPS Data Base. U.S. Environmental Protection Agency (EPA-600/4-80-013a).

Ruff, R. E. 1983. Application of statistical methods to diagnose causes of poor air-quality model performance. Atmos. Environ., 17(2):291-297.

Ruff, R. E., H. S. Javitz, and J. S. Irwin. 1980. "Development and Application of a Statistical Methodology to Evaluate the Realtime Air Quality Model (RAM)." Second Joint Conference on Applications of Air Pollution Meteorology. American Meteorological Society.

SAS. 1985. SAS User's Guide: Basics, Version 5. SAS Institute, Inc., Cary, North Carolina.

Scire, J. S., and L. L. Schulman. 1981. "Evaluation of the BLP and ISC Models with SF_6 Tracer Data and SO_2 Measurements at Aluminum Reduction Plants." APCA Specialty Conference on Dispersion Modeling for Complex Sources, St. Louis, Missouri.

Singer, I. A. 1961. The relationship between mean and peak concentration. J. Air Pollut. Control Assoc., 11(7):336-341.

Smith, M. E. 1984. Review of the attributes and performance of 10 rural diffusion models. Bull. Amer. Meteorol. Soc., 65(6):554-558.

Stull, R. B. 1983. Integral scales for the nocturnal boundary layer. Part 1: Empirical depth relationships. J. Climate Appl. Meteor., 22:673-686.

Texas Air Control Board. 1979. "User's Guide to the Texas Episodic Models." Austin, Texas.

Touma, J. S., Y.-H. Huang, H. Wang, and J. H. Christiansen. 1981. "Development and Validation of a Multi-Source Plume Downwash Model." Preprints: Twelfth International Technical Meeting on Air Pollution Modeling and Its Application. North Atlantic Treaty Organization, Committee on the Challenges of Modern Society.

Turner, D. B. 1970. Workbook of Atmospheric Dispersion Estimates. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Turner, D. B., and J. S. Irwin. 1983. "Comparison of Sulfur Dioxide Estimates from the Model RAM with St. Louis RAPS Measurements." In Air Pollution Modeling and Its Application II, C. De Wispelaere; ed., Plenum, New York.

Turner, D. B., and J. S. Irwin. 1982. Extreme value statistics related to performance of a standard air quality simulation model using data at seven power plant sites. Atmos. Environ., 16(8):1907-1914.

Turner, D. B. 1985. Proposed pragmatic methods for estimating plume rise and plume penetration through atmospheric layers. Atmos. Environ., 19(7):1215-1218.

Turner, D. B., J. S. Irwin, and A. D. Busse. 1985. Comparison of RAM model estimates with 1976 St. Louis RAPS measurements of sulfur dioxide. Atmos. Environ., 19(1):247-253.

Venkatram, A. and R. Paine. 1984. "Model to Estimate Dispersion of Elevated Releases into a Shear-Dominated Boundary Layer." Fourth Joint Conference on Applications of Air Pollution Meteorology." American Meteorological Society, Boston, Massachusetts.

Weil, J. C. 1979. Applicability of stability classification schemes and associated parameters to dispersion of tall stack plumes in Maryland. Atmos. Environ., 13:819-831.

Weil, J. C., and R. P. Brower. 1982. "The Maryland PPSP Dispersion Model for Tall Stacks." Martin Marietta Corporation, Baltimore, Maryland.

Weil, J. C., and R. P. Brower. 1984. An updated Gaussian plume model for tall stacks. J. Air Pollut. Control Assoc., 34:818-827.

White, W. H., C. Seigneur, D. W. Heinold, M. W. Eltgroth, L. W. Richards, P. T. Roberts, P. S. Bhardwaja, W. D. Conner, and W. E. Wilson, Jr. 1985. Predicting the visibility of chimney plumes: An intercomparison of four models with observations at a well-controlled power plant. Atmos. Environ., 19(3):515-528.

Willis, G. E., and J. W. Deardorff. 1978. A laboratory study of dispersion from an elevated source within a modeled convective planetary boundary layer. Atmos. Environ., 12:1305-1311.

Willis, G. E., and J. W. Deardorff. 1987. Buoyant plume dispersion and inversion entrapment in and above a laboratory mixed layer. Atmos. Environ., 21(8):1725-1735.

APPENDIX

Maximum One Hour SO_2 Concentrations Analyzed by Scaling Procedures

by

Steven R. Hanna
Sigma Research Corporation
394 Lowell St., Suite 12, Lexington, MA 02173

1. Introduction

One hour SO_2 concentrations are the basis for most SO_2 air quality data collection efforts and are implicit in the predictions of most atmospheric dispersion models. In this short note, maximum one hour observed SO_2 concentrations are compared with the predictions of a few simple dispersion models for maximum concentrations. It is recognized that maximum concentrations from very tall stacks ($h_s > 150m$) occur during light-wind convective conditions and maximum concentrations from moderately tall stacks ($50m < h_s < 100m$) occur during high wind conditions. The stack height marking the transition between these two regimes will be calculated below.

2. Convective Scaling

Hanna and Paine (1987) use SF_6 data from the Kincaid and Bull Run power plants to verify a simple scaling law for maximum one-hour ground level concentrations anywhere in the vicinity of a power plant in convective conditions:

$$C_{*c} = 0.04/F_* \qquad (1)$$

where $C_{*c} = Cuh^2/Q$
$F_* = F/uw_*^2(h-h_s)$

and the formula is valid only for $F_* > 0.06$. This restriction assures that the plume is sufficiently buoyant to rise through the mixed layer. Equation (1) has the dimensional form:

$$C = 0.04\ (Q/F)w_*^2(h-h_s)/h^2 \qquad (2)$$

Because the buoyancy flux at a power plant is usually proportional to the emissions of SO_2, the ratio Q/F for SO_2 should be a constant at a given power plant (Figure 1 suggests that $Q/F \approx 4\ gs^2\ m^{-4}$ at Kincaid). The absolute maximum of Equation (2) occurs when $h = 4h_s$, yielding

$$C_{max} = 0.0075\ (Q/F)w_*^2/h_s \qquad (3)$$

Preceding page blank

This simple scaling formula will be compared with similar formulas for high wind maxima below.

3. High Wind Scaling

Maximum concentrations occur during high winds for stack heights in the range from 50 to 100m. Plume rise is very low during these conditions and the bottom of the plume diffuses quickly to the ground. Formulas for overall maximum hourly concentrations for a one year sampling period for these conditions are given by Hanna, Briggs and Hosker (1982):

$$C_{*hw} = C_{max}\, uh_s{}^2/Q = 0.092 \tag{4}$$

The results of a special study of high wind maximum concentrations at the Baldwin and Clifty Creek power plants yield $C_{*hw} = 0.051$ and 0.059, respectively. For the purpose of this analysis, it will be assumed that $C_{*hw} = 0.075$, which is midway between the theoretical and the observed values.

4. Moderate Wind Daytime Scaling

The two formulas given above do not cover daytime situations with moderate winds when the $F_* > 0.06$ criterion is not satisfied. In this case the Weil et al. (1986) pdf model can be used:

$$C_* = 0.4C_{y*}/\sigma_{y*} \tag{5}$$

$$\sigma_{y*} = 0.56X_*/(1+0.7X_*)^{1/2} \tag{6}$$

$$C_{y*} = (0.048/\sigma_{z1*})\exp(-h_{1*}{}^2/2\sigma_{z1}{}^{*2}) + (0.036/\sigma_{z2*})\exp(-h_{2*}{}^2/2\sigma_{z2*}{}^2) \tag{7}$$

where $X^* = xw_*/uh$ $\qquad h_{s*} = h_s/h$

$C_{y*} = C_y\, uh/Q$

$\sigma_{z1*}{}^2 = 0.21F_*{}^{2/3}X_*{}^{4/3} + 0.24\, X_*{}^2$

$\sigma_{Z2*}{}^2 = 0.21F_*{}^{2/3}X_*{}^{4/3} + 0.48\, X_*{}^2$

$h_{1*} = h_{s*} + 1.6F_*{}^{1/3}X_*{}^{2/3} - 0.35X_*$

$h_{2*} = h_{s*} + 1.6F_*{}^{1/3}X_*{}^{2/3} - 0.40X_*$

This is called the pdf model because it explicitly accounts for the non-Gaussian pdf of the vertical velocity distribution in convective conditions.

The concentration predicted by equation (5) is the hourly-averaged ground level plume centerline value. To obtain the maximum value for that period, the solution at several X_* must be calculated. Resulting values of C_{max*} as a function of h_{s*} and F_* are given in Table 1. The absolute maximum dimensional C that would result from this method is not clearly evident from this table containing dimensionless C_*. It is evident, however, that the values in Table 1 agree fairly well with the high wind model (Equation 4 with a constant of 0.075) if it is assumed that $F_*/h_{s*} = 0.1$. This result is consistent with the result quoted in Hanna, Briggs, and Hosker (1982) that the critical wind speed and the maximum hourly concentration occur when plume rise Δh equals one third of the stack height, where

$$\Delta h = 1.54\ (F/uu_*^2)^{2/3}\ h_s^{1/3} \qquad (8)$$

With this condition, the maximum dimensional concentrations occur for $F_*/h_{s*} = 0.1$ in Table 1, and it is concluded that the pdf model is consistent with the high-wind model at the point of maximum concentration.

5. Comparison of High Wind and Convective Maxima

At any given site, the ratio of the highest hourly averaged maximum concentrations for strongly buoyant plumes during high-wind convective conditions is given by the formula:

$$C_{maxhw}/C_{maxc} = 10F_c/(u_{hw}w_{c*}^2h_s) \qquad (9)$$

where subscript hw and c refer to high-wind and convective conditions, respectively. The wind speed u_{hw} at which the maximum C_{maxhw} occurs is given by Hanna, Briggs, and Hosker (1982) as $u_{hw} = 13.1\ (F/h_s)^{1/3}$, yielding the following result for the critical stack height, h_s', at which the high-wind and convective maxima are equal for a given F and w_*:

$$h_s' = 0.66\ F/w_*^3 \qquad (10)$$

For $h_s > h_s'$ the convective maximum dominates, and for $h_s < h_s'$ the high-wind maximum dominates.

Recognizing the definition $w_* = (H_oh)^{1/3}$ and the finding that $h=4h_s$ at the time of the absolute convective maximum, this equation can be written in the form:

$$h_s' = 0.41\ (F/H_o)^{1/2} \qquad (11)$$

Thus the critical stack height h_s' is a weak (square root) function of the ratio of the plume buoyancy flux to the ambient heat flux. At the Kincaid and Bull Run sites, F ranges from 500 to 1000 m^4s^{-3}. Briggs (1975) finds that the ambient heat flux, $H_o = (g/T)\ \overline{w'T'}$, is about 0.02 m^2s^{-3}. Consequently, h_s' is calculated to equal 65m for F equal to 500 m^4s^{-3} and 130m for F equal to 2000m^4s^{-3}. Since the actual stack height at Kincaid and Bull Run are 187m and 203m, respectively, it is expected that the convective maximum will dominate at those sites. An analysis of a data set containing 5000 hours of one hour average maximum SO_2 concentrations at Kincaid verifies this expectation, for the observed convective maximum (487 ppb) is over two times the observed high-wind maximum (203 ppb).

6. Discussion of Equation (11)

Equation (11), which expresses the critical stack height separating dominance of the convective maximum from dominance of the high wind maximum, is the major "product" of this analysis. It certainly is intuitively obvious that, the larger the ambient heat flux H_o, the more likely a given stack will show dominance of the convective maximum. However, it is somewhat counter-intuitive that, the larger the buoyancy flux from the stack, the less likely a given stack will show dominance of the convective maximum. This paradox is the result of a further condition on the formula for the convective maximum:

$$F^* = F/uw_*^2(h-h_s) > 0.06 \qquad (12)$$

This condition assures that the plume travels to the top of the mixed layer and is consequently partly imbedded in the capping inversion. The larger F, the more of the plume is lost to the capping inversion, and the smaller the convective maximum will be. If the condition in Equation (12) is not met at a given stack, equation (11) is no longer valid. In this case, the pdf model applies to unstable conditions, and since it was shown in section 4 that the pdf model maximum is equivalent to the high-wind model maximum, it is concluded that the site is dominated by the high-wind maximum.

References:

Briggs, G.A., 1975: Plume Rise Predictions. Lectures on Air Pollution and Environmental Impact Analyses. AMS, Boston, 59-111.

Hanna, S.R., G.A. Briggs and R.P. Hosker, 1982: Handbook on Atmospheric Diffusion. DOE/TIC 11223, Tech. Inf. Center, USDOE, Oak Ridge, TN.

Hanna, S.R. and R.J. Paine, 1987: Convective Scaling Applied to Diffusion of Plumes from Tall Stacks. To appear in Atm. Environ.

Weil, J.C., L.A. Corio and R.P. Brower, 1986: Dispersion of Buoyant Plumes in the Convective Boundary Layer. Proc. Fifth Joint Conf. on Appl. of Air Poll. Meteorol., AMS, Boston, 335-338.

Table 1

Maximum Concentrations, $C_{max*} = C_{max}\,uh^2/Q$, as a Function of h_{s*} and F_* for the PDF Model

F_* h_{s*}	0	0.001	0.005	0.010	0.020	0.050
0.01	3530	703	245	134	67.6	24.3
0.02	889	245	103	61.8	33.9	13.5
0.05	144	55.8	19.9	19.4	12.0	5.60
0.10	36.7	17.4	10.3	7.45	5.03	2.64
0.20	9.51	5.29	3.50	2.71	1.97	1.16
0.50	1.67	1.01	0.82	0.68	0.54	0.36

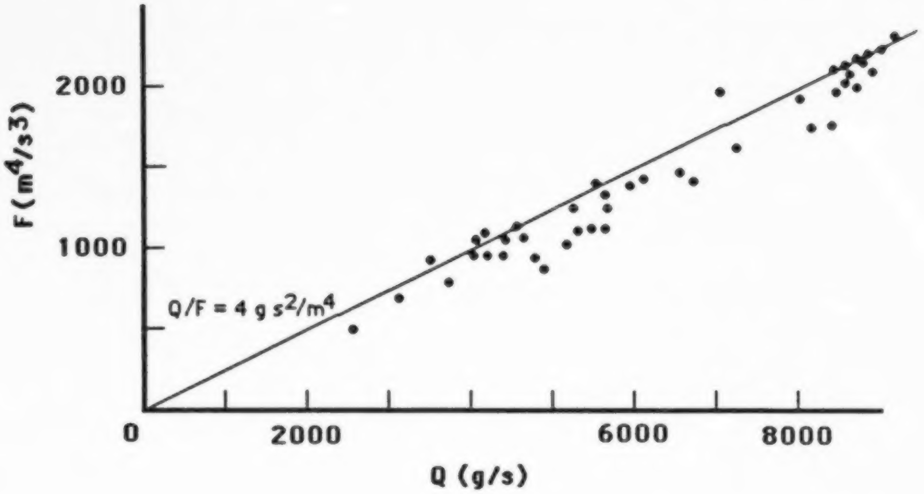


FIGURE 1. Kincaid--top 25 SO_2 hours for DEV and EVAL data bases.